A method for identifying and evaluating energy meter data based on big data analysis technology

Chencheng Wang

A method for identifying and evaluating energy meter data based on big data analysis technology

Chencheng Wang

State Grid Sichuan Electric Power Company Marketing Service Center,
Renmin S. Road, Wuhou District Chengdu, Sichuan, 610000, China
Email: 3099523591@qq.com

Abstract: In order to explore the measurement performance of grid energy meters under multi-dimensional influence conditions on site and map their measurement errors under standard laboratory conditions, a measurement error estimation method for on-site service energy meters based on big data analysis technology is proposed, which combines environmental data and electrical factor data from on-site operation to achieve online measurement error estimation. To address the problem of electricity meter demand prediction, a reasonable optimisation allocation model for electricity meters based on Shapley combination model and neural network is established to improve the accuracy of demand prediction. By mining historical data, Holt Winters, BP neural network, and RBF neural network models are used to predict, compare, and analyse the demand for electricity meters. The test results indicate that the built model can achieve reliability evaluation based on the real-time operating status of intelligent energy meters, providing auxiliary decision-making for the operation and maintenance of intelligent energy meters.

Keywords: intelligent energy meter; electric energy error; temperature; power factor; BP neural network.

Biographical notes: Chencheng Wang holds a Master's degree and is a Senior Engineer. His main research field is electricity metering and informatisation.

1   Introduction

Smart energy meters are measuring instruments that are subject to mandatory national verification and management. Their measurement errors not only affect the interests of thousands of households, but also affect the safety, stability, and economic operation of smart grids. Currently, under laboratory reference conditions (Hu et al., 2023; Huo et al., 2023), error verification is usually carried out on electricity meters that have expired before grid installation and eight years of rotation. This method cannot perform measurement error verification on electricity meters that are in grid operation. In addition, portable devices can also be used for on-site calibration of energy meters, or the meters can be disassembled for laboratory calibration. However, considering the large number of smart energy meters, with hundreds of millions of grid connected energy

meters, whether it is dismantling the meters back to the laboratory for calibration or using portable equipment for on-site calibration, it requires a lot of manpower and material resources, a large workload, high cost, high labour intensity, and low work efficiency. The measurement error of smart energy meters is seriously affected by environmental conditions. The actual operating conditions of on-site grid operation are complex and random, and there is a problem of inconsistency between operating conditions and laboratory reference conditions. This leads to the risk of exceeding the standard for laboratory certified energy meters during on-site operation. Therefore, the evaluation of on-site operation measurement errors of electric energy meters under multidimensional time-varying conditions has become a difficult problem that needs to be urgently solved in this field.

At present, the following three methods are mainly used to establish a mathematical model for reasonably evaluating the reliability of intelligent energy meters. Firstly, by establishing a failure rate model using the component stress method, it is possible to effectively simulate faults or abnormal situations of smart energy meters based on hardware structure and functional design, and calculate failure – mean time to failure (MTTF) based on the component manual to characterise the estimated lifespan of smart energy meters, It can be used as an important parameter for reliability evaluation of intelligent energy meters (González-Cagigal et al., 2021; Andrade et al., 2020; Chen et al., 2023); Secondly, based on a pre assumed distribution of the failure efficiency of an intelligent energy meter, such as the Wilson distribution, the least squares fitting of the parameters is performed on the fault data of the intelligent energy meter to obtain a reliability evaluation model for the intelligent energy meter (Rasooli and Itard, 2020); in addition, combined with the assumed failure efficiency distribution of smart energy meters and the multi stress Peck model, an accelerated life test was conducted on smart energy meters to evaluate their reliability (Zezheng et al., 2022; Guorui et al., 2019; Air China et al., 2023). However, current research mostly faces the problem of highly simplified functional topology of smart energy meters and static parameters of the formed life model, which cannot reflect or adapt to the impact of changes in the physical model and operating environment conditions of smart energy meters (Lu et al., 2023; Yang et al., 2022a; Yang et al., 2022b; Wang et al., 2022b). The accelerated life test is more complex, and as the reliability level of smart energy meters gradually improves, the cost of obtaining failure sample data required in the test is relatively high.

With the continuous improvement of business systems such as the power user electricity information collection system (referred to as the collection system) and marketing business system (Zhang et al., 2022), the synchronisation and completeness of power data collection continue to improve. The use of power big data to achieve online monitoring of the operating status of electricity meters has become a new effective control method. The existing online monitoring methods mainly include energy balance in the substation area, clustering of electricity consumption status, and building evaluation systems. The principle of energy balance is the foundation of most research on measurement inaccuracy analysis. Wang and Ouyang (2022) constructs equations based on the low-voltage substation assessment table and the electricity balance relationship of user meters, and uses linear regression algorithm to solve and obtain the operating errors of each energy meter; Wen et al. (2022) defines the actual flow increment of a generalised flow instrument using reading increment and relative error under the principle of flow conservation, and derives an equation for the reading increment and relative error of the instrument. In state clustering, Ding et al. (2022) uses a BP neural

network classifier to monitor the operating error of the electricity meter; Lu et al. (2023) focuses on clustering features of electricity consumption behaviour in both vertical time and horizontal space, achieving large-scale anomaly detection of electricity consumption data streams. In the evaluation system evaluation, Beibei (2022) divided the energy metering device into four parts: energy meter, current transformer, voltage transformer, and secondary circuit. Combined with operating status data, the operating errors of each equipment were evaluated; Yu et al. (2022) calculates the failure rate, scrap cost, and depreciation rate of electricity meters based on 'manufacturer + batch'.

To address the shortcomings of existing reliability evaluation methods, a modelling method for intelligent energy meter reliability evaluation based on multi-source data fusion is proposed for reasonable evaluation of intelligent energy meter reliability. After integrating and organising multi-source big data of smart energy meters, survival analysis theory is used to model them, and they are used as covariates that affect the reliability of smart energy meters to characterise their survival function. At the same time, a deep neural network was used to learn the survival function parameters, and a survival function model of intelligent energy meters under the influence of covariates was obtained. The evaluation results were compared with traditional methods such as classical component stress method through example analysis, verifying the rationality and feasibility of the model. According to the analysis of existing literature, the research on demand forecasting of electric energy meters can be mainly divided into single model forecasting and combined model forecasting. Compared with the single model prediction, the combined model prediction can further improve the prediction accuracy by combining each single model and selecting the optimal weight to optimise the results. The core of combination model prediction is how to choose the appropriate combination method. The Shapley method (Wang et al., 2022a) is a method proposed by Shapley in 1953 to handle the allocation of benefits for the participation of various alliance members in the overall goal. It can effectively handle the contribution allocation problem of each participant in cooperation and has been well applied in power demand forecasting and other aspects. The article uses the Shapley combination method to model and predict the demand for electricity meters. Three models, Holt Winters (Feng et al., 2022), BP neural network, and RBF neural network, were compared and combined. The Shapley combination method was used to allocate weights to the predicted results of each model, in order to exert the predictive effect of each model and achieve comprehensive prediction results. Using the LM optimised BP neural network data training method, combined with load current, power factor and other data, a mapping relationship model between measurement error and environmental temperature, humidity, load current, and power factor stress is constructed, that is, a multi-dimensional measurement error model suitable for grid connected operation of electric energy meters is established; Then, by learning and modelling different types of environmental data, the estimation accuracy of energy meter error under on-site operating conditions is improved, and the measurement error of the grid energy meter under laboratory conditions is derived; Finally, the effectiveness and rationality of the proposed method are verified by using the field data validation of an area.

2   Method and principle

2.1   Design of energy error model for intelligent energy meters

The improved on-site environmental data K-means (Dai et al., 2022) clustering analysis flowchart is shown in Figure 1. In terms of constructing the power error model, the designed neural network structure is a multi-layer feedback network structure (Zhu et al., 2022), with full interconnection between layers and no interconnection between the same layer. Moreover, the BP neural network structure includes the input layer, hidden layer and output layer, which is a neural network based on the back-propagation training algorithm of model error. The three-layer nonlinear network can approach any continuous function with any precision, and has good generalisation ability (Liu et al., 2022b). Therefore, this paper selects a typical three-layer BP neural network topology structure, that is, the hidden layer is a single layer, The structural diagram of the neural network designed in this article is shown in Figure 2.

This article clusters the on-site environmental temperature and humidity, combines the load current and load power factor, and uses BP neural network (Fan et al., 2022) to estimate the measurement error of electric energy meters operating on the network. Considering that the units of these four variables are different and the order of magnitude is also significant, before using BP neural network to establish an electric energy meter measurement error model, the load current and load power factor are also classified and processed, Assign values to different load currents and load power factors separately, and try to ensure that the assignment range is between [0, 1]. Specifically, when the load current is 0.05Ib, 0.1Ib, 0.5Ib, Ib, and Imax, the values are assigned as 0.05, 0.1, 0.5, 0.8, and 1, respectively; Load power factor is 1, 0.5 L.

Assign values of 1, 0.5, and 0.8 respectively at 0.8 C. Research shows that temperature, humidity, load current and power factor (Senave et al., 2019) have a serious impact on the metering error of electric energy meters. Therefore, the homogenisation values of these four stresses are selected as the input parameters of the neural network model, that is, the number of neurons in the model input layer $a = 4$, and the input temperature and humidity values are the environmental temperature and humidity values in a collection after clustering the on-site environmental cluster analysis (Liu et al., 2022a). After passing through the input layer neurons (Chen, 2022), the on-site environmental data and electrical factor data are the inputs of each neuron in the hidden layer.

$$H_f = \sum_{f=1}^{4} w_{fg} r_f - \beta_g \tag{1}$$

Among: w_{fg} Set the weight from the input layer to the hidden layer; β_g. The threshold from the input layer to the hidden layer. $f = 1, 2, \ldots, a, g = 1, 2, \ldots, b$.

B is the number of neurons in the hidden layer; A is the number of neurons in the input layer and is 4. And the number of hidden layer neurons

$$b = \sqrt{a+c} + a_1 \tag{2}$$

Figure 1   K-Means cluster analysis flowchart (see online version for colours)
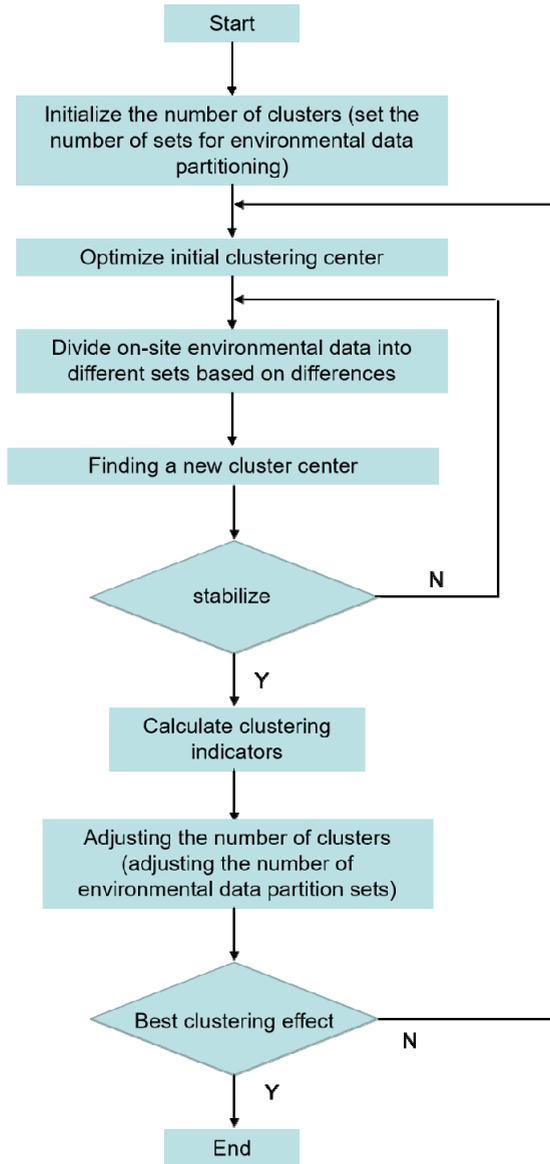
In the formula: c is the number of neurons in the output layer; $A1$ is a constant and $1 < a1 < 10$.

After obtaining information, the hidden layer processes the data and transmits it to the output layer. Considering that the input ambient temperature value, humidity value, load current value and load voltage value are normalised, this paper sets the activation function as an S-type function $f(x) = \dfrac{1}{1+e^{-x}}$. Its output is limited between (0, 1). Since

the positive number of the activation function is chosen as an S-type function, the output of each node in the hidden layer is

$$H_g = f\left(\sum_{f=1}^{4} w_{fg} r_f - \beta_g \right) \tag{3}$$

Figure 2 BP neural network structure (see online version for colours)
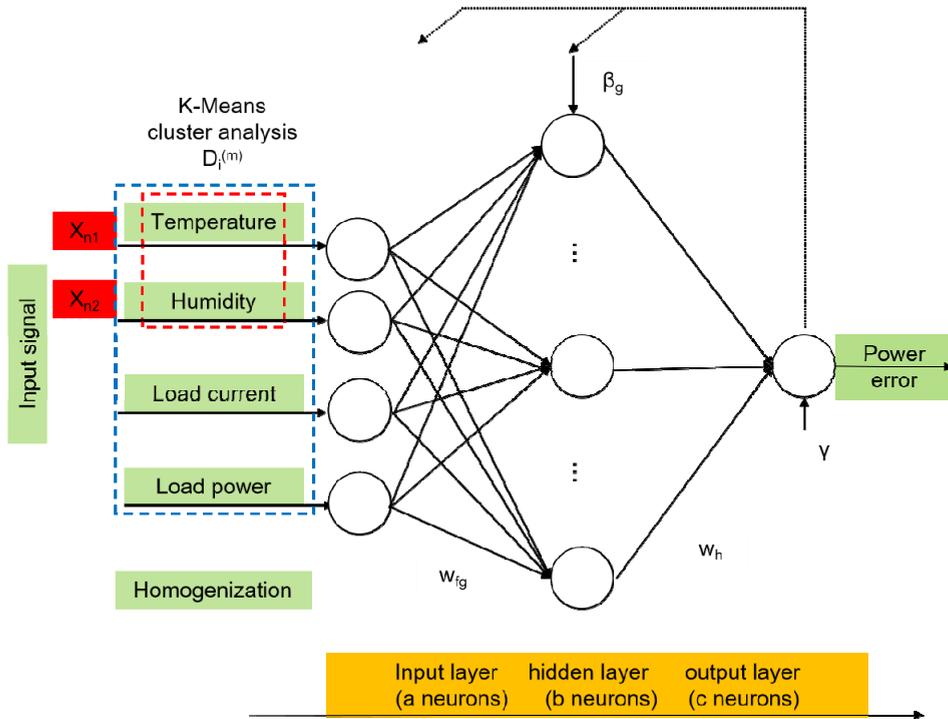
The output signal of the model built by the BP neural network is the energy error of the electricity meter, so the number of neurons in the output layer of the model is $c = 1$. Let the threshold from the hidden layer to the output layer be γ. The weight from the hidden layer to the output layer is wh ($h = 1, 2, …, b$), then the input signal of the output layer is

$$Y_f = w_h H_g - \gamma \tag{4}$$

Since the activation function is an S-type function, the output function of the output layer can be obtained as

$$y = f\left(w_h H_g - \gamma \right) \tag{5}$$

Compare the estimated error value with the actual error value, and when there is a significant difference between the two, the difference will be transmitted back from the hidden layer to the input layer. That is, the learning process of the energy meter error follows the forward transmission of information and the reverse transmission of learning differences. The neural network model continuously adjusts its weight and difference threshold, ultimately achieving a difference in output parameters (Li et al., 2021) that is

lower than the pre-set learning accuracy. Among them, traditional BP neural networks have problems such as slow convergence speed, and may even not converge, and are prone to falling into local minima. Many scholars have proposed relevant improvement algorithms. This article has decided to use the LM (Zhang et al., 2021) algorithm to improve the shortcomings of traditional BP networks, improve the convergence speed of the network, and increase the accuracy of network training. The training error function and the corresponding Jacobian matrix are obtained, and then the weights of the corresponding nodes are adjusted to Wang et al. (2017) using the LM optimisation algorithm

$$\Delta w = \left(J^T J + \mu I \right)^{-1} \cdot J^T e \tag{6}$$

In the formula, J is the Jacobian matrix of the weight differentiation of errors; E is the error vector; μ is a scalar.

2.2 *Online monitoring method for operating status of low-voltage side energy meters*

In the article, the operation status of the low-voltage side energy meter is monitored online through component comparison method. The sub component comparison method combines the neighbouring voltage comparison method and the zero line current analysis method to evaluate the operating status of the metering chip, voltage sampling component, and current sampling component of the energy meter, reducing the estimation of the energy meter status from the meter level to the component level (Xu et al., 2017; Zheng et al., 2016), providing an effective online monitoring method for the low voltage side operating status.

   The specific steps for comparing components are as follows:

1 By querying the event records, characteristic voltage, and response denial of the energy meter in the substation area through the collection system, determine whether there is a communication failure between the main CPU and the metering chip. If there is no fault, proceed to step (2)

2 Collect voltage, load curve, live line current, zero line current, phase, and clock data from electricity meters at the same time section through a data acquisition system, and perform data pre-processing (Liu, 2012)

3 Perform voltage data pre-processing, combined with neighbouring voltage comparison method, to analyse the voltage of the same box and phase electric energy meter, and locate the abnormal voltage of the electric energy meter under the substation area

4 Perform current data pre-processing and locate abnormal current meters under the substation area using zero live line current analysis method

5 Analyse the correlation between the data obtained from steps (3) and (4) and the type of component fault: if there is a voltage anomaly and no current anomaly, it is determined that the voltage sampling component is faulty; If there is a current anomaly and no voltage anomaly, it is determined that the current sampling component is faulty; If there are abnormalities in both, due to the low possibility of

simultaneous failure of voltage sampling and current sampling components, it is judged that the reference voltage component of the metering chip is faulty.

Figure 3 Flow chart of component comparison method (see online version for colours)
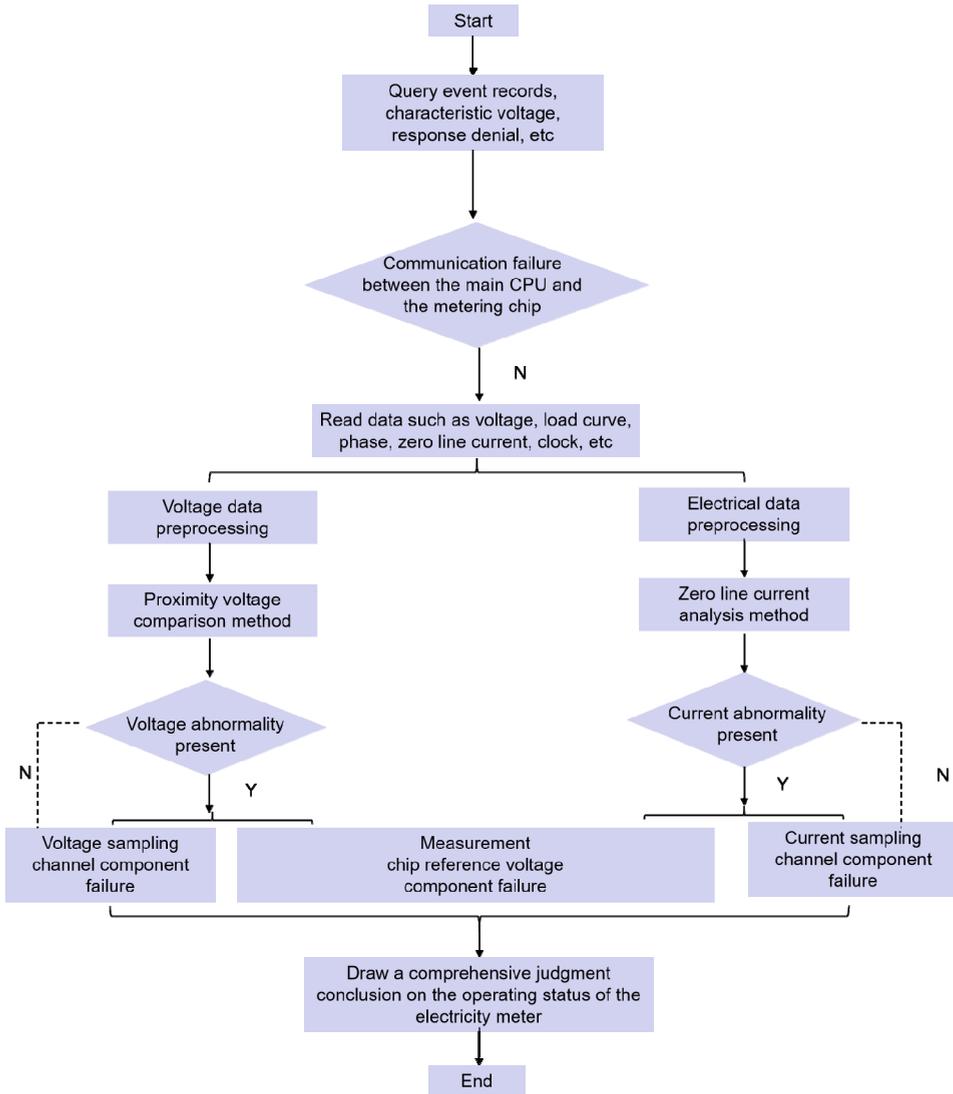
The flowchart of the component comparison method is shown in Figure 3.

In step (3), the neighbouring voltage comparison method is used. As the voltage measurement values of the same meter box and the same phase of the meter tend to be consistent, the voltage measurement values of the monitored meter are close to those of other meters in the same box and phase. Therefore, a linear regression model for the neighbouring voltage comparison at zero crossing is selected. Determine the voltage measurement error of the electric energy meter by performing linear regression (Wang

and Ouyang (2022; Wen et al., 2022; Ding et al., 2022) with the voltage curve of the same box and same phase electric energy meter, as shown in equations (7) and (8)

$$y_u = \beta_u x_u \tag{7}$$

$$\varepsilon_u = (1 - \beta_u) / \beta_u \cdot 100\% \tag{8}$$

In the formula, x_u is the voltage measurement value of adjacent reference energy meter A; Y_u is the voltage measurement value of the energy meter B that needs to be compared; βU is the estimated value of linear regression parameters for voltage data; εU is the voltage measurement error of the energy meter B; Using x_u as the independent variable and yu as the dependent variable, perform linear regression when accumulating enough measurement samples β_u. Thus, the voltage measurement error of the energy meter can be calculated ε_u (Zhu et al., 2009).

Step (9): The electric energy meter with abnormal current is located using the zero line current analysis method. In the standardised connection method of intelligent electric energy meters, the electric energy meter and the load are connected in series through the live line and zero line, so the zero line current measurement values tend to be consistent. Using the zero line current of a single-phase meter as a reference standard (Dong, 2007), construct a linear regression model for zero and live line analysis of zero crossing. The model is as follows:

$$y_1 = \beta_I x_I \tag{9}$$

In the formula, x_I is the measured value of the zero line current as the independent variable; Y_I is the measured value of the live wire current as the dependent variable. When accumulating enough samples, solve through linear regression β_I parameter estimate. Current measurement error ε_I is

$$\varepsilon_I = (1 - \beta_I) / \beta_I \cdot 100\% \tag{10}$$

Evaluate the operation status of the metering chip, voltage sampling element (Gao et al., 2006), and current sampling element of the energy meter, reduce the estimation of the energy meter state from the meter level to the element level, and provide an effective online monitoring method for the low voltage side operation status.

The article focuses on the operation status of energy meters for high supply and low metering enterprise users. On the medium voltage side, with complete power supply data at each measuring point on the line, regression analysis can be used to locate inaccurate energy meters using line side energy balance. However, in the actual application process, the measurement points on the line side and user side are managed by different business systems, and there are issues with data synchronisation and integrity, as well as professional barriers in business control. Therefore, the article proposes an online monitoring method for the operation status of the medium voltage side based on the external characteristics of the transformer, which only uses the user side measurement data to locate the inaccurate energy meter, and has strong practicality (Zhang et al., 2004).

The external characteristics of a transformer refer to the obvious pattern of the voltage on the low-voltage side of the transformer changing with the load. Equivalent the

transformer to a T-shaped circuit (Wen et al., 2022), with *Rm* and *Xm* as excitation impedances (Zhang, 2004); *Rh* and *Xh* are the leakage impedances on the high-voltage side; *Rl* and *Xl* are the leakage impedances on the low-voltage side. Due to the much smaller excitation current than the user's normal load current, the impact on the line voltage drop is very small. *Rm* and *Xm* are ignored, and *Rt* and *Xt* are used to represent the sum of the corresponding impedances *Rh, Xh, Rl*, and *Xl*, respectively. Given the transformer capacity, *Rt* and *Xt* can be inferred based on the transformer model parameters. Set the high voltage side voltage to *Ea, Eb, Ec*, and combine it with on-site wiring to obtain the three-phase equivalent circuit for high supply and low supply special transformer users, as shown in Figure 4.

Figure 4 Three-phase equivalent circuit for high supply and low metering special transformer users
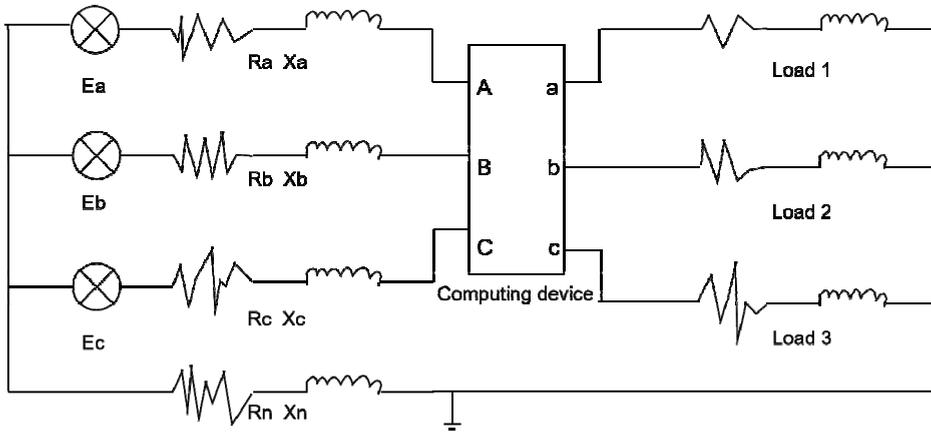
2.3 Holt winters model

The Holt Winters model, also known as the cubic exponential smoothing model, adds a seasonal term on top of the horizontal term in the primary exponential smoothing method and the trend term in the secondary exponential smoothing method. The Holt Winters model includes two types: addition model and multiplication model. The selection of the two models mainly depends on whether the trend of the data varies seasonally by addition or multiplication. The addition Holt Winters model data is overlaid in addition form, and the formula is as follows:

$$u_i = \alpha(x_i - S_{i-T}) + (1-\alpha)(u_{i-1} + v_{i-1}) \tag{11}$$

$$v_i = \beta(u_i - u_{i-1}) + (1-\beta)v_{i-1} \tag{12}$$

$$S_i = \gamma(x_i - u_i) + (1-\gamma)S_{i-T} \tag{13}$$

In the equation α, β, γ. The smoothing coefficients for the horizontal term, trend term, and seasonal term are all within the range of [0, 1], which are the balance weights between the predicted values of the model and the measured inverse values; S_i is the exponential smoothing value of the *i*-th season term, and T is the length of the season

cycle; Vi is the exponential smoothing value of the *i*-th trend term; U_i is the exponential smoothing value for the *i*-th period. The prediction formula for *xi* is the actual value of the *i*-th period, as follows

$$\hat{x}_{i+h} = u_i + hv_i + S_{i-T+h} \qquad (14)$$

Among, $\hat{x}_{i+h}$ is the predicted value for the $i + h$ period, h is the number of backward smoothing periods, and $h > 0$.

The network structure of BP neural network mainly includes three parts: input layer, hidden layer, and output layer. If the number of input neurons in the network is M and the number of output neurons is N, then the neural network structure can be regarded as a mapping from an M-dimensional Euclidean space to an N-dimensional Euclidean space. The training process of BP neural network is to send the values of the previous layer to the next layer through weighted averaging, and then transmit them to the output layer. The resulting error is then transmitted to the previous layer, and finally the weighted average is recalculated. After repeated training, the error is reduced until the set error is met.

The steps of BP neural network prediction are as follows:

1   Import monthly fault energy meter data and determine the number of nodes. Reasonably divide the given historical data, determine the input and output parts of the data, and then determine the number of input neurons and output neurons

2   Determine the number of hidden layer neurons.

The selection of the number of hidden layer neurons has a significant impact on the prediction performance of BP neural networks. The number of hidden layer neurons should be selected within an appropriate range. If there are too many neurons, it will cause excessive data fitting and increase the computational workload; A small number of neurons can lead to poor prediction performance. The number of hidden layer nodes in BP neural networks is not only affected by the number of input and output nodes, but also by the actual complexity of the problem itself and the expected error set. When selecting the number of neurons in the hidden layer, the number of neurons in the hidden layer should be 2–3 times that of the input layer or selected based on empirical formulas in literature, as follows:

$$L = \sqrt{M + N} + A \qquad (15)$$

In the formula, *M* is the number of input layer nodes; *N* is the number of output layer nodes; *L* is the number of hidden layer neurons; *A* is a constant between [1, 10];

3   Determine the training function. Use MATLAB's neural network toolbox for network training.

In 1985, Powell proposed the radial basis function (RBF) for multivariate interpolation. The RBF neural network consists of three parts: input layer, intermediate layer, and output layer. The first layer is the input layer composed of the input signal source, which is a linear function; The second layer is the hidden layer, which is a nonlinear function. The number of hidden elements in the hidden layer is determined by the model problem; The third layer is the output layer, which is a linear function and responds to the input signal. The RBF neural network model is shown in Figure 5.

Figure 5 RBF neural network model diagram
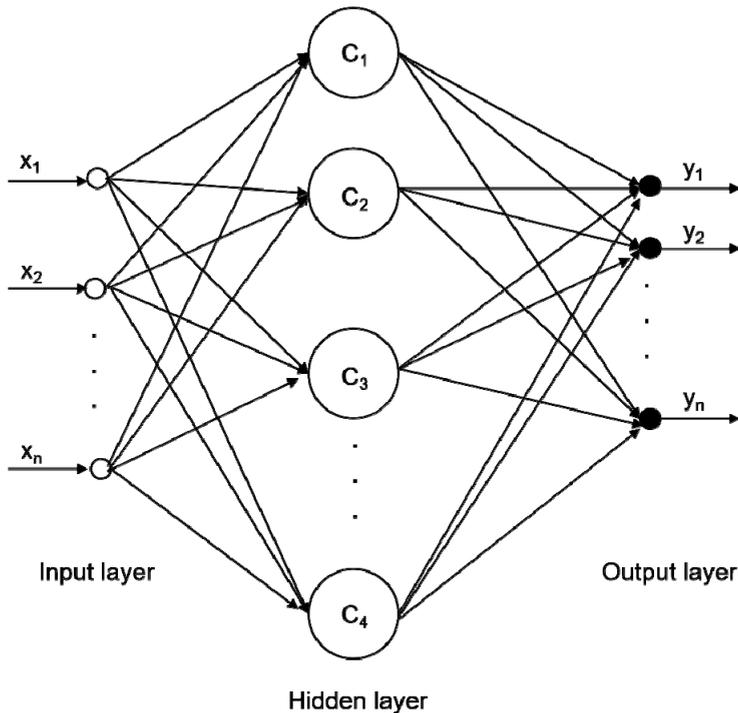
RBF neural network sends linear input vectors to the hidden layer, uses the hidden layer to transfer data from low-dimensional space to high-dimensional space, making it linearly separable in high-dimensional space, and finally sends it to the output layer through linear weighting. RBF neural network is a feedforward neural network that, due to its linear output and relatively easy parameter adjustment, does not have local optimisation problems. Therefore, RBF neural networks have advantages such as optimal approximation and global optimality.

3 Experiment

3.1 Example of reliability evaluation for intelligent energy meters

Based on the actual operation and maintenance data of smart energy meters in a certain city, the effectiveness of the reliability evaluation model for smart energy meters proposed above was verified. The training samples in the experiment contain a total of 49,640 maintenance data and abnormal data of smart energy meters. The dataset is divided into a training set Dtrain and a testing set Dtest in a 4:1 ratio, with the number of abnormal types $N = 21$. Hardware platform for the experiment: The operating system is Windows 8.1, the CPU is Core single core i5-5200 U, 2.20 GHz, and the code implementation is based on Python's lifelines library package and TFDeepSurv library package. Input the training model shown in Figure 6 for training.

Figure 6   Model training flowchart
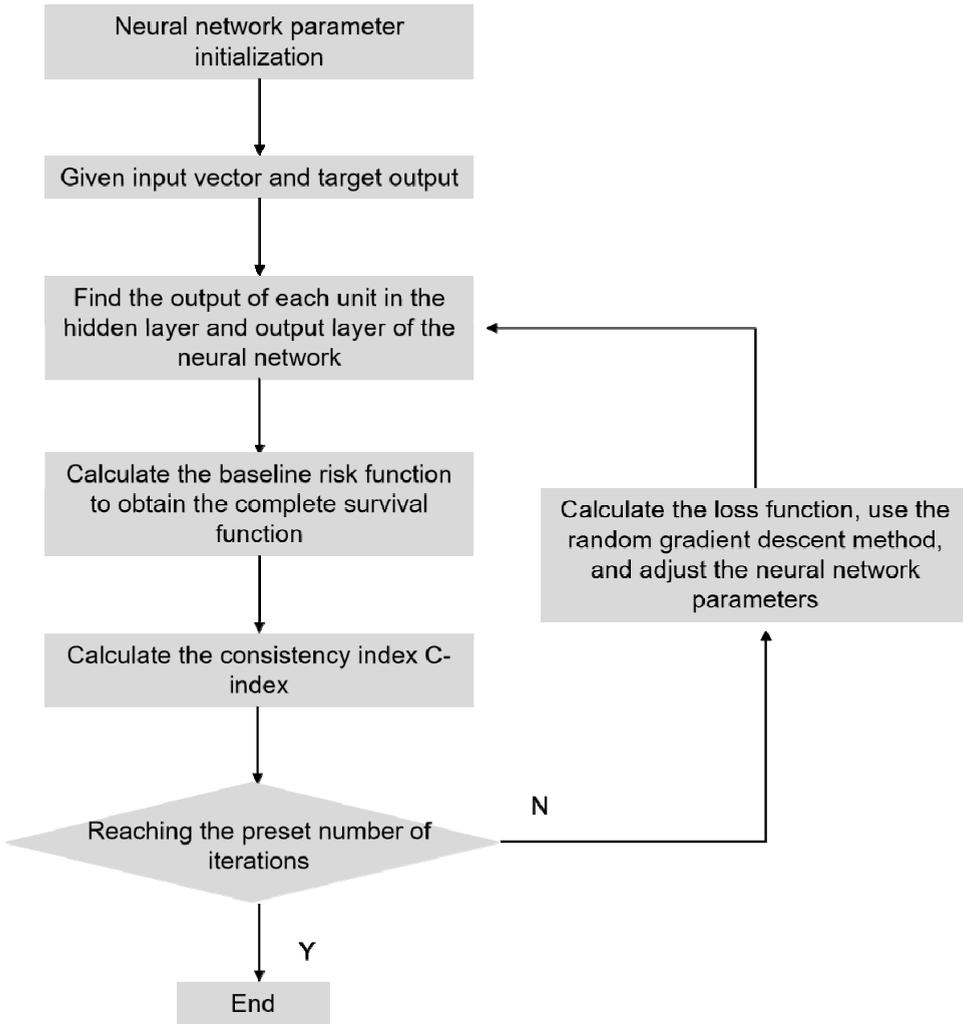
The change trend of loss function during training is shown in Figure 7. The consistency index obtained on DTRAIN is about 0.682, and the consistency index obtained on Dtest is about 0.683.

To represent the survival curves of the reliability evaluation model obtained under different covariate values, the covariate vectors X (1), X (2), X (3) are taken to investigate the impact of any anomaly corresponding to the covariate, such as anomaly 5, satisfying the constraints shown in equation (15)

$$x_5^{(1)} + 0.2 = x_5^{(2)} + 0.1 = x_5^{(3)} \tag{15}$$

The covariate components corresponding to other types of anomalies in X (1), X (2), and X (3) have the same values. When the covariate vectors obtained from the integration of multi-source big data of three different smart energy meters are X (1), X (2), and X (3), their corresponding survival probability curves are shown in Figure 7. It can be seen that

as the value of $x(i)$ 5 increases, the survival probability of the corresponding intelligent energy meter individual i at the same time decreases, the reliability of the intelligent energy meter significantly decreases, and the likelihood of failure or replacement according to established strategies increases. The obtained reliability evaluation model, as an effective reliability evaluation model, can reflect the real-time health status of smart energy meters. On this basis, power grid operators establish an optimal operation and maintenance model for smart energy meters by comprehensively considering factors such as economic benefits of power grid operation and maintenance and replacement costs, and set input parameters according to work requirements to obtain the optimal strategy for smart energy meter operation and maintenance.

Figure 7 Change of loss function during training (see online version for colours)
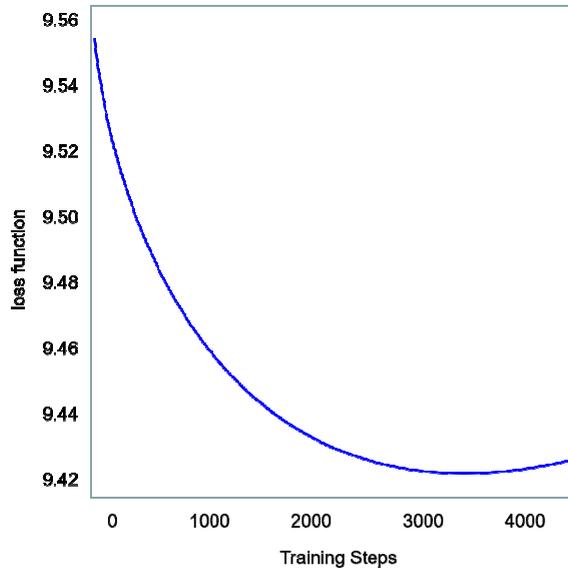
In order to compare the advantages and disadvantages of different prediction methods, based on the relevant data of the same batch of smart energy meters, the component stress method was used to calculate their mean life before failure (MTTF), and the prior assumed reliability curve distribution function was used for parameter fitting to evaluate the reliability of smart energy meters, and compared with the established evaluation model. To evaluate the reliability of smart energy meters using the component stress method, a list of smart energy meter components needs to be obtained. Intelligent energy meters are generally considered as a simple series failure model, namely

$$\lambda_s = \sum_{i=1}^{N} \lambda_i \tag{16}$$

Among λ_s Is the system failure rate (1/h), which is the failure rate of~N components (1/h). For a certain model of smart energy meter, by analysing the list of components on the smart energy meter motherboard and selecting JB/Z299C-2006 electronic equipment prediction manual, the system failure of the smart energy meter is calculated as $\lambda\ S =$ 7.881672 (10-6/h), then the average time before failure (MTTF) of the smart energy

meter is $1/\lambda$ S = 14.48 years. Assuming that the time failure rate relationship of smart energy meters follows a Weibull distribution, based on existing smart energy meter data, the maximum likelihood estimation method is used to obtain the reliability curve of the smart energy meter, as shown in Figure 9.

Figure 8 Example of survival curves with different covariate values (see online version for colours)
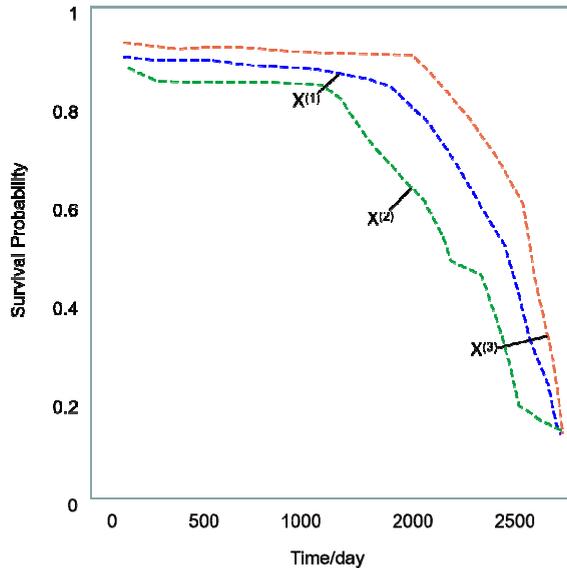
The above three methods have effectively evaluated the reliability of smart energy meters from three different perspectives, each with different criteria for judging their strengths and weaknesses, and cannot be uniformly measured using concepts such as accuracy and C – index. Therefore, comparisons are made from two aspects: the model's generalisation ability and the interpretability of the results. In terms of generalisation ability, the component stress method needs to search for the corresponding failure rates in the component list and component manual of each type of smart energy meter. The resulting model changes with the replacement of smart energy meter components or component composition, resulting in poor generalisation ability. The parameter fitting method based on hypothesis distribution and the prediction method proposed in the article both rely on the selected intelligent energy meter fault data samples for training, and the generalisation ability of the model depends more on the sample quality. In terms of interpretability of prediction results, the component stress method can analyse the failure of key components inside smart energy meters, and establish a series relationship between micro component failure and functional failure. However, when a smart electricity meter malfunctions, it often manifests as the failure of a certain functional module, making it difficult to locate a specific component, and the obtained MTTF is a certain value, which has limited guiding significance for the operation and maintenance of smart electricity meters. However, assuming the reliability distribution of smart energy meters in advance and using data to fit model parameters using empirical assumptions cannot correspond to the endogenous or exogenous reliability impact factors of smart

energy meters, resulting in poor interpretability of the evaluation curve. The reliability evaluation model established in the article, which considers the influence of multiple covariates, can be strengthened and maintained by analysing the impact of a single covariate on the survival probability curve.

Figure 9 Survival function based on Weibull distribution fitting (see online version for colours)
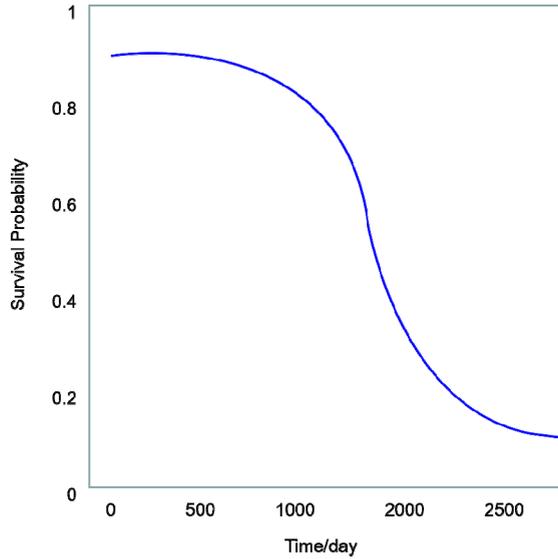
Figure 10 Distribution of energy meters in the meter box
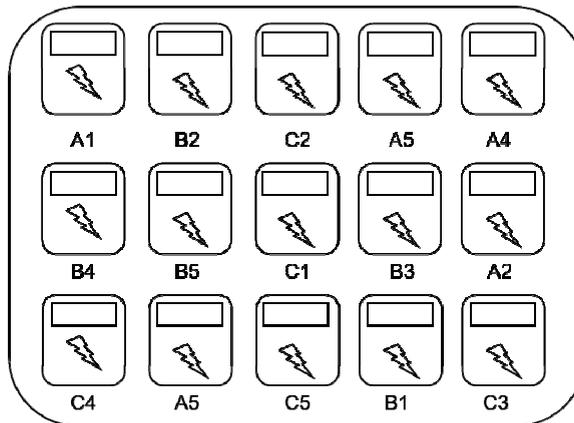
3.2 *Example analysis*

To verify the effectiveness of the method proposed in the article in solving the operating states of different levels of electricity meters, multiple instances were modelled and solved. The computer used is Inter (R) Core (TM) CPU I5-3470 3.20 GHz, 4 GB of memory, and programming languages are Python and GAMS23.9

Use the component comparison method in Part 1 to analyse the actual meter box in a certain community. This meter box has a total of 15 A, B, and C three-phase energy meters, with 5 for each phase. The distribution diagram inside the meter box is shown in Figure 10.

Figure 11   Linear regression results between user # A5 and user # A1
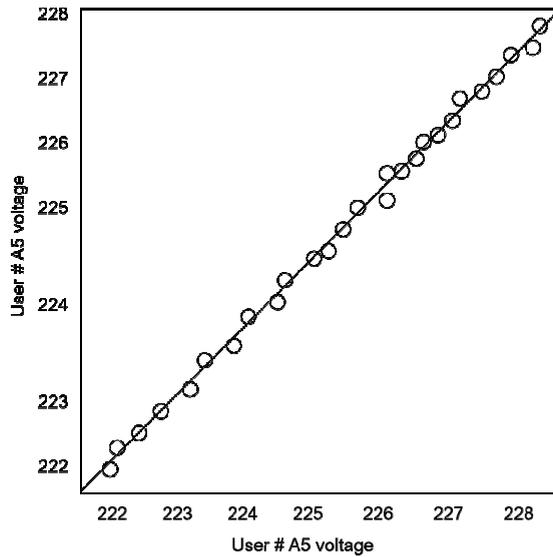
According to the analysis of the main station of the collection system, the communication between the main CPU of the energy meter under the meter box and the metering chip is good and in normal operation. Using the neighbouring voltage comparison method, the voltage of the same box and same phase electric energy meter in the substation area is respectively substituted into equation (1) for linear regression, and the zero live line current of the electric energy meter in the box is substituted into equation (3) for linear regression. The single-phase user # A5 is located with current anomalies and no voltage anomalies, and it is determined that there is metering error in the current sampling channel component. The calculation process is as follows. Take the 96 point voltage curve of 5 A-phase energy meters in the meter box as input samples.

Perform linear regression between user # A5 voltage amplitude and other energy meters in the same box and phase. Taking # A1 as an example, the results of linear regression are shown in Figure 11.

By analogy, perform linear regression on the voltage curves of users # A5 and users # A1~# A4 in the same box, and the calculation results of the regression coefficients are shown in Table 1. Analysis shows that the voltage component of user # A5 electricity meter is in normal metering state.

Table 1   Calculation results of adjacent voltage comparison method for user # A5

	#A1	#A2	#A3	#A4
#A5	1.0002	0.9998	1.0001	1.0002
εu	–0.02%	0.02%	–0.01%	0.01%

Similarly, the zero line current analysis method was used to perform linear regression on 10 samples collected on the same day and at the same time. The regression coefficient was calculated to be 0.9623, and the current measurement error was 3.92%. The accuracy level corresponding to the 220 V single-phase electrical energy metering device is required to be 2% as the allowable error range of the energy meter (Feng et al., 2022). Therefore, the current measurement error of 3.92% exceeds the allowable error range of the energy meter. Based on the above analysis, there is a measurement inaccuracy in the current sampling element of the electricity meter corresponding to user # A.

3.3 Shapley combination model

This section discusses the prediction methods based on the Shapley combination model mentioned above.

Analyse the effectiveness. This data is presented in Yang et al. (2022a) and Based on the data, obtained through sample expansion. The original data series of monthly fault demand for single-phase meters used from 2017 to 2022 are shown in Table 2.

Table 2   Impact of the number of hidden layer nodes on the prediction results of BP neural network

Number of hidden layer nodes	8	9	10	11	12
actual error	0.00123	0. 00122	0.00163	0. 00142	0. 00132
Maximum mean square error (mse)	$2.70 \times 10 - 4$	$6. 08 \times 10 - 4$	$7. 30 \times 10 - 4$	$4. 69 \times 10 - 4$	$4. 53 \times 10 - 4$

Figure 12   Prediction effect of Holt Winters model (see online version for colours)
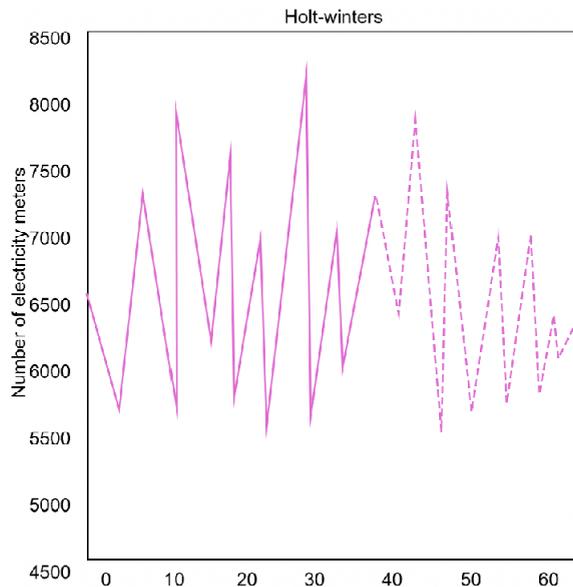
Prepare monthly single-phase electricity meter data from 2017 to 2022.

Forecast the monthly demand for single-phase electricity meters in 2022. The fitting prediction effect of the Holt Winters model is shown in Figure 12, where the abscissa represents the data for each month of 60 months; The vertical axis represents the number of electricity meters required. By comparing the predicted results of the Holt Winters model with the actual data of single-phase electricity meters in 2022, the prediction error of the Holt Winters model can be calculated.

The calculation method for prediction error is: prediction error=| predicted value – actual value |/actual value × 100%.

Compile monthly single-phase fault energy meter data from 2017 to 2022.

Train for the training set. Select the data from the first four years to predict the data for the fifth year, and use the 2022 data as the test set for comparison. In the BP neural network, four neurons are set in the input layer and one neuron in the output layer, with a training frequency of 20 times and a performance index error of 0.002. Select default values for other parameters (weights and thresholds). Randomly select the initial parameters used to train the BP neural network. Use BP neural network to fit the curve, find the relationship between input and output values, and iteratively optimise until one of the preset conditions is met to stop training. The impact of the number of nodes in the hidden layer on the prediction results of the BP neural network is shown in Table 2. Randomly select the initial parameters used to train the BP neural network. Use BP neural network to fit the curve, find the relationship between input and output values, and iteratively optimise until one of the preset conditions is met to stop training. The impact of the number of nodes in the hidden layer on the prediction results of the BP neural network is shown in Table 2. Among them, mse is the mean square error, which is the expected value of the square of the difference between the estimated value of the parameter and the true value of the parameter. According to Table 2, in order to reach a compromise between overfitting and prediction accuracy, the paper selects 11 hidden layer nodes.

4　Conclusions

Based on the multi-source big data of smart energy meters, a reliability evaluation model for smart energy meters was established by integrating and analysing the maintenance data and abnormal data of smart energy meters. The CoxPH model combined with deep learning was used to fit the multi-source data for fusion analysis, and a lifespan survival probability model for intelligent energy meters was obtained. The Holt Winters model, BP neural network model, and RBF neural network model were used to fit and predict the monthly electricity meter demand, and compared with the actual value of the monthly electricity meter. Then, the Shapley combination model method was used for combination modelling. Due to the large prediction deviation of the Holt Winters model, a combination model was constructed using BP neural network and RBF neural network, and the demand for electricity meters was predicted based on the combination model. Verified the advantages of fast convergence and high accuracy of the proposed method; And finally, the conversion relationship curve between the measurement error of the electric energy meter under on-site working conditions and the measurement error under laboratory reference conditions was provided. The measurement error of the electric energy meter was divided and the electric energy meters with larger measurement errors were selected, which helps to improve the error verification work of the electric energy

meters operating on the grid and can predict the out of tolerance failure of the measuring equipment in on-site operation in advance, Helps to improve the inspection efficiency of metering errors in electric energy meters operating on the grid.

References

Air China, Jing, R., Xiaoli, H., Cui, Q., Wang, X. and Kong, Y. (2023) 'Research on prediction of type 2 diabetes nephropathy based on optimized neural network', *Health Vocational Education*, Vol. 41, No. 1, pp.155–159, DOI: 10.20037/j.issn.1671-1246.2023.01.48.

Andrade, S., Alves, J., Barbosa, J. et al. (2020) 'Smart outlet: smart electrical outlet with device identification using NFC', *Journal of Autonomous Intelligence*, Vol. 2, No. 4, p.21.

Beibei, S. (2022) 'Backstepping control of flexible manipulators based on neural networks', *University Physics Experiment*, Vol. 35, No. 6, pp.106–111.

Chen, G., Hu, X., Hong, X., Dang, S. and Zhou, L. (2023) 'Subsidence deformation monitoring using Kalman filtering and wavelet neural network', *Geospatial Information*, Vol. 21, No. 1, pp.101–103 + 112.

Chen, P. (2022) *Research on the Recognition of Rotational Indication of Electric Energy Meters in Complex Environments*, Guangdong University of Technology.

Dai, Y., Dong, X., Liu, Y., Liang, Y., Liu, X., Qi, J. and Sun, Y. (2022) 'Method for estimating measurement errors of electric energy meters in field operation under big data conditions', *Journal of Harbin Institute of Technology*, Vol. 27, No. 5, pp.97–105.

Ding, J., Zhou, Y., Yuan, T., Xia, J., Li, W. and Han, Y. (2022) 'Research progress of anesthesia related neural networks in monitoring anesthesia depth', *Chinese Journal of Clinical Pharmacology and Theapeutics*, Vol. 27, No. 12, pp.1400–1407.

Dong, H. (2007) *Energy Meter Calibration and Information Management System Based on BP Neural Network,* Harbin Institute of Technology.

Fan, Y., Deng, X., Gao, Y., Tang, Z., Zhu, J. and Zhu, L. (2022) 'Improving PSO optimization of RBF intelligent energy meter terminal temperature detection method', *Journal of Electric Power Science and Technology*, Vol. 37, No. 5, pp.207–214.

Feng, X., Guo, B., Chen, J., He, S. and Zhao, Y. (2022) 'Research on the construction and implementation method of a quality evaluation model for electric energy meters', *Application of Microcomputers*, Vol. 38, No. 12, pp.85–88 + 97.

Gao, Z., Dong, H. and Song, W. (2006) 'A verification method for electric energy meters based on Bp neural network', *Electrical Measurement and Instrumentation*, Vol. 2006, No. 9, pp.37–39.

González-Cagigal, M.A., Rosendo-Macías, J.A. and Gómez-Expósito, A. (2021) 'Application of nonlinear Kalman filters to the identification of customer phase connection in distribution grids', *International Journal of Electrical Power and Energy Systems*, Vol. 2021, No. 125, p.106410.

Guorui, W.U., Jiayu, S.U., Dixing, L.I. et al. (2019) 'Anti-RF interference design of RFID meters', *Automation and Instrumentation*, Vol. 13, No. 3, pp.113–122.

Hu, J., Ying, G., Qing, T. and Ying, Z. (2023) 'Hardware implementation of neural network accelerator based on RISC-V', *Electronics and Packaging*, Vol. 23, No. 2, pp.72–77, DOI: 10.16257/j.cnki.1681-1070.2023.0016.

Huo, B., Gao, Y. and Qi, X. (2023) 'Research on pulse neural networks for brain like computing', *Chinese Journal of Image and Graphics*, Vol. 28, No. 2, pp.401–417.

Jin, Y., Xu, Y. and Huang, X. (2021) 'Research on the confidence interval of energy meter measurements based on convolutional neural networks', *Chinese Journal of Electrical Engineering*, Vol. 41, No. 19, pp.6630–6641.

Li, C., Shen, H., Liu, J., Wu, Y., Sun, X. and Zhang, Y. (2021) 'Research on electricity meter demand prediction based on Shapley combination model and neural network', *Electrical Measurement and Instrumentation*, Vol. 58, No. 9, pp.187–193.

Liu, C. (2012) *Research and Application Of Energy Consumption Analysis Model For Large Public Buildings Based On Genetic BP Neural Network Technology*, Guangxi University.

Liu, W., Liu, L., Liu, X., Cui, H. and Li, Y. (2022a) 'Evaluation model for error state of intelligent energy meters in low voltage distribution stations based on improved BP neural network', *Electrical Measurement and Instrumentation*, Vol. 59, No. 11, pp.176–181.

Liu, Y., Lu, X., Li, R., Yang, G. and Li, N. (2022b) 'Research on fault diagnosis method for gateway measurement devices using RBF neural networks', *Journal of Electrical Engineering*, Vol. 2022, No. 20, pp.133–136 + 139.

Lu, Y., Zhou, S. and Hu, X. (2023) 'Prediction of solar radiant intensity intensity based on BP neural network and wavelet neural network', *Software Engineering*, Vol. 26, No. 1, pp.5–8 + 4, DOI: 10.19644/j.cnki.issn2096-1472.2023.001.002.

Rasooli, A. and Itard, L. (2020) 'Automated in-situ determination of buildings' global thermo-physical characteristics and air change rates through inverse modelling of smart meter and air temperature data', *Energy and Buildings*, Vol. 229, No. 2, pp.57–69.

Senave, M., Roels, S., Verbeke, S. et al. (2019) 'Sensitivity of characterizing the heat loss coefficient through on-board monitoring: a case study analysis', *Energies*, Vol. 12, No. 17, p.3322.

Wang, X., Cao, M., Jiang, X., Zhao, Y., Li, X., Zhao, X., Jiang, T. and Yuan, Y. (2017) 'Measurement error prediction and correction of gateway electric energy metering devices based on BP neural network', *Journal of Yunnan University for Nationalities (Natural Science Edition)*, Vol. 26, No. 6, pp.497–501.

Wang, X., Ouyang, P. (2022) 'Overview of the application of artificial neural networks in the field of concrete structures', *Low Temperature Building Technology*, Vol. 44, No. 12, pp. 75–78.

Wang, X., Zhu, E. and Ai, Z. (2022a) 'The principle of convolutional neural networks and their application in medical imaging diagnosis', *Chinese Journal of Medical Physics*, Vol. 39, No. 12, pp.1485–1489.

Wang, Y., Guo, M., Jiang, J., Li, M. and Wang, S. (2022b) 'Application of BP neural network model algorithm in settlement observation of a certain building', *Jingwei Tiandi*, Vol. 2022, No. 6, pp.57–60 + 63.

Wen, S., Ma, Y. and Sun, Z. (2022) 'Short term prediction of photovoltaic power generation based on GWO-EEMD-BP neural network', *Journal of Central South University (Natural Science Edition)*, Vol. 53, No. 12, pp.4799–4808.

Xu, J., Li, C. and Li, Y. (2017) 'Electricity theft analysis based on genetic algorithm optimized BP neural network', *Software*, Vol. 38, No. 11, pp.18–23.

Yang, B., Ruan, J., Li, J. and Chen, Z. (2022a) 'Research on product production date recognition technology based on residual neural networks', *Science and Technology Wind,* Vol. 2022, No. 36, pp.73–75.

Yang, C., Zhang, S., Xu, C. and Hao, R. (2022b) 'Membrane distillation performance prediction system based on BP neural network', *Zhejiang Chemical Industry*, Vol. 53, No. 12, pp.34–37 + 46.

Yu, S., Yin, S., Xue, R. and Li, Y. (2022) 'Application of MEA-BP neural network in predicting earthquake simulation test data', *Journal of Gansu Science*, Vol. 34, No. 6, pp.31–7, DOI: 10.16468/j.cnki. issn1004-0366.2022.06.006.

Zezheng, X., Yuan, Z., He, X., Fang, S., Wang, Q., Qi, X. and Chen, X. (2022) 'Application of local linear wavelet neural network in atomic clock signal prediction', *Journal of Time and Frequency*, Vol. 45, No. 4, pp.279–285.

Zhang, C., Zhu, Y., Han, L. and Xia, Y. (2022) 'Research on damage localization of grid structures based on RBF neural network', *Journal of Shaoyang University (Natural Science Edition)*, Vol. 19, No. 6, pp.41–48.

Zhang, J., Wu, J., Ou, X. and Ou, X. (2021) 'Automatic recognition of foreign object sound in electric energy meters based on feature reduction dimension neural network', *Mechanical Design and Manufacturing*, Vol. 2021, No. 3, pp.234–237.

Zhang, P. (2004) *A Harmonic Energy Measurement Error Analysis Device Based on BP Neural Network,* Wuhan University.

Zhang, P., Xiang, T., Zhou, Q. and Chen, H. (2004) 'Harmonic measurement error analysis software based on BP neural network', *High Voltage Technology*, Vol. 2004, No. 10, pp.62–63, DOI: 10.13336/j.1003-6520.hve.204.10.026.

Zheng, M., Li, X., Ju, H., Pang, F. and Li, S. (2016) 'Research on fault classification of energy meter software based on BP neural network', *North China Electric Power Technology*, Vol. 2016, No. 8, pp.8–12, DOI: 10.16308/j.cnki. issn1003-9171.2016.08.002.

Zhu, Z., Chen, H., Jiang, C., Zhen, H., Xu, Y. and Tong, T. (2022) 'Research on the measurement inaccuracy of a type of voltage sampling resistance anomaly in electric energy meters', *Journal of Central South University for Nationalities (Natural Science Edition)*, Vol. 41, No. 6, pp.682–688.

Zhu, Z., Ge, W. and Cui, B. (2009) 'Research on remote energy measurement system based on BP neural network', *Volkswagen Business*, Vol. 2009, No. 24, p.292.