Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 663–672, DOI 10.12694/scpe.v26i2.3957

QUALITY ANALYSIS AND PREDICTION METHOD OF SMART ENERGY METER BASED ON DATA FUSION

SIWEI WANG[*], JI XIAO[†], YINGYING CHENG[‡], YU SU[§], AND WENLI CHEN[¶]

Abstract. In order to study the quality analysis method of key links in smart energy meters, the author proposes a data fusion based quality analysis and prediction method for smart energy meters. This method is based on the relevant data of key links in the electric energy meter, and selects the data of the electric energy meter in research and development design, material procurement, production and manufacturing, acceptance testing, installation and operation, dismantling and scrapping as the sample data for model construction. The XGBoost algorithm classification method is used to establish an intelligent electric energy meter quality analysis model. Taking the dismantled electricity meter data of a certain power company as an example, this paper conducts modeling analysis and prediction of various quality issues of smart electricity meters, and conducts on-site verification. Based on the verification results, the model is continuously optimized. The results indicate that: The model was optimized using cross validation and grid search methods, and the final model achieved an accuracy rate of 0.74 and a recall rate of 0.82 on the validation set. This method can meet the actual needs of power grid business and objectively reflect the quality situation of key links in smart energy meters.

Key words: Energy meter failure rate, Time series, Fault characteristics, XGBoost algorithm, multiple linear regression

1. Introduction. With the development of human living standards and society, more and more power electronic components are being applied to the power system. With the integration of national photovoltaic poverty alleviation projects, the proportion of distributed photovoltaic grid connection is increasing. These devices have relatively superior performance, but they have caused an impact on the power quality, making the problem of power quality in the low-voltage platform area increasingly severe. How to comprehensively monitor and evaluate the power quality of low-voltage substation areas has become one of the hot topics for power supply enterprises and researchers [1].

Excellent power quality is an important guarantee for the safe and economic operation of the power grid. Power quality issues not only cause losses to electricity consuming enterprises and customers, but also seriously affect the power supply service indicators of power supply enterprises. Even serious power quality problems will impact the brand image of power supply enterprises [2]. If there is a voltage quality issue, excessive voltage can cause damage to equipment such as transformers, energy meters, and electrical appliances used by customers; Low voltage can bring huge obstacles to social production and human life, and serious low voltage problems may cause machines to malfunction and cause economic losses [3]. For example, the three-phase imbalance problem in the low-voltage substation area can slightly reduce the efficiency of low-voltage lines and distribution transformers, but in severe cases, it may cause serious consequences such as wire overload burning, switch burning, and even single-phase burning of distribution transformers [4]. How to carry out power quality monitoring in low-voltage substations is the primary issue in analyzing power quality issues. The power quality monitoring device can provide real-time data monitoring for power supply enterprise staff, and assist them in recording and analyzing the basic situation of power quality in low-voltage substation areas. The monitored operational data can also be used to analyze the problems of electricity customers and provide effective solutions for grassroots grid managers [5].

[*]State Grid Chongqing Electric Power Company Marketing Service Center, Chongqing, 400023, China. (Corresponding author, SiweiWang5@163.com)
[†]State Grid Chongqing Electric Power Company Marketing Service Center, Chongqing, 400023, China. (JiXiao13@126.com)
[‡]State Grid Chongqing Electric Power Company Marketing Service Center, Chongqing, 400023, China. (YingyingCheng65@163.com)
[§]State Grid Chongqing Electric Power Company Marketing Service Center, Chongqing, 400023, China. (YuSu32@126.com)
[¶]State Grid Chongqing Electric Power Company Marketing Service Center, Chongqing, 400023, China. (WenliChen59@163.com)

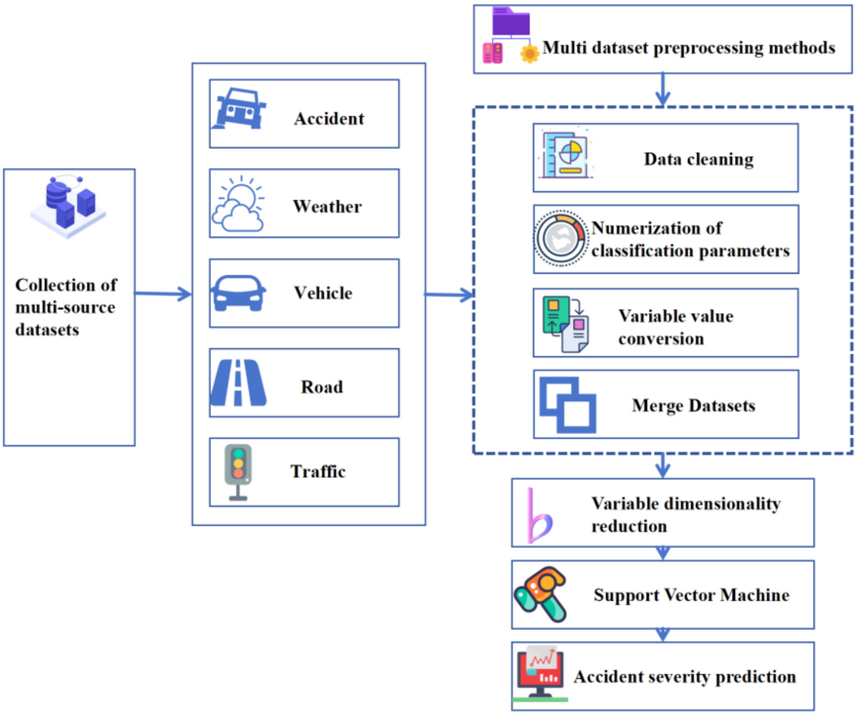

Fig. 1.1: Quality Analysis of Intelligent Energy Meters Based on Data Fusion

The power quality monitoring device can also record the type and geographical location of faults in power supply equipment, which helps to carry out power supply repair services and improve the efficiency of restoring power supply. With the deployment of digital transformation strategies for power supply enterprises, more and more perception devices are being applied to power supply, and a plethora of new technologies (big data, cloud platforms) are gradually being applied to various fields of power supply enterprises. Like the right wing front flag of Horqin, the power supply service resource scheduling and control system covers the entire area, and the perception ability of power grid equipment has been greatly improved [6]. The promotion and construction of business systems such as electricity information collection, online power grid, and power supply service resource scheduling and control provide strong data support for comprehensive monitoring of the operation status of the power grid. According to the author's statistics, all provincial companies of State Grid Corporation of China have established data service platforms and massive data platforms, but progress in data value mining, data analysis applications, data cleaning and integration is relatively slow [7] (Figure 1.1).

2. Literature Review. The price of power quality monitoring devices is expensive, and considering their cost, it is not possible to configure them in large quantities in the distribution network. Therefore, the optimization goal of minimizing power quality monitoring points has always been a research hotspot for power quality monitoring schemes. The goals of power quality monitoring are different, and the methods of configuring power quality monitoring points are also different, making it difficult to form a unified standard. Himeur, Y. proposed a monitoring device configuration scheme that takes into account the severity of voltage sag in substations, taking into account the number and observability of substation monitoring, and taking into account the number of monitoring points and the observability of voltage sag at each node of the entire network as constraints [8]; Chen, Y. proposed an equipment configuration optimization scheme for monitoring the entire network voltage using the fault point method, taking into account the number of monitoring points and the observability of voltage dips at each node, in response to voltage dips caused by line short circuits [9]; Nakutis, Ž. proposed in [10] a method of using particle swarm optimization algorithm to optimize equipment configuration for monitoring points with voltage sag; Spertino, F. proposed a monitoring point configuration algorithm using an improved particle swarm optimization algorithm by reasonably setting the minimum number of monitoring points [11]; Karngala, A. K. proposed an equipment configuration optimization plan that takes into account the severity of voltage sag in substations, addressing the issue of existing power quality monitoring point layout schemes not taking into account the type, management requirements, and equipment configuration sequence of each monitoring point [12].

Electricity information collection data is usually stored as historical data in historical databases, and some scholars have begun to explore the value of this data; Zhou, M. analyzed electricity information collection data and proposed an electricity theft identification method based on electricity feature analysis, which can be used to screen suspected electricity theft users [13]; Ma, J. studied the fast clustering and anomaly detection techniques for power data flow in large-scale power information collection, and designed and implemented a flow clustering algorithm based on the clustering characteristics of power behavior in vertical and horizontal spaces, achieving fast clustering and anomaly detection [14,15].

The author reviews the quality related data of key links in electric energy meters and studies the method of extracting quality impact features; Compare various big data analysis technologies and establish a quality analysis model for smart energy meters; Use this model to predict and analyze potential quality hazards of smart energy meters, and conduct on-site verification. Continuously optimize the model based on the verification results.

3. Research Methods. The task of the electricity meter quality analysis model is to mine the patterns of faults in dismantled electricity meters based on relevant data of key links of electricity meters, predict the probability of faults in operating electricity meters with the same characteristics, and conduct on-site data verification.

The research and development design, material procurement, production and manufacturing, acceptance testing, installation and operation, and dismantling and scrapping processes that have a significant impact on the quality of electric energy meters are defined as key links. The data situation of each link is sorted out to facilitate subsequent data selection [16].

For key link data, use Pearson correlation coefficient and chi square test to conduct correlation analysis on data fields. Based on the threshold reference given by business experts, delete some fields with correlation coefficients greater than 0.5, and finally use the key link data of the electricity meter. After cleaning and transformation, generate data that can be used for modeling and analysis. The sample data of the model training set is based on historical data from Henan. In the data selection stage, a total of 130 fields were selected from the original data.

Analyze the 132 original features based on the key links of the electricity meter according to the following steps:

The first step is data visualization. In order to visually present the relationship between the characteristics and whether the electricity meter is faulty, these 132 original features were used to draw the distribution graphs of the faulty electricity meter and the normal electricity meter in each feature [17]. The distribution of fault table and normal table on several typical features is shown in Figure 3.1. As shown in the figure, there is no significant difference in the distribution of faults in each feature of the energy meter, and further feature extraction is needed through quantitative indicators.

The second step is to select features based on the Gini impurity method. The calculation formula is shown in equation 3.1.

$$I_G(f) = \sum_{i=1}^{m} f_i(1-f_i) = \sum_{i=1}^{m} f_i - f_i^2 = 1 - \sum_{i=1}^{m} f_i^2 \tag{3.1}$$

In the formula, m represents the total number of categories; f_i is the probability that the sample points belong to class i.

Calculate the Gini importance of each feature by taking the reciprocal of Gini impurity, as shown in equation 3.2, and the results are shown in Table 3.1.

$$gini = \frac{1}{I_G(f)} \tag{3.2}$$

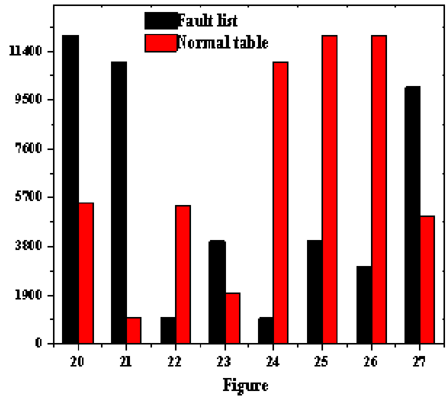

Fig. 3.1: Comparison of distribution of fault table and normal table on various features

Table 3.1: Feature Importance (Partial)

Name	Characteristic	Importance
The number of days the electric energy meter ran during the first abnormal collection	DAYS_BEFORE_FIRST_FA	122
The region code of the payment terminal	AREA_CODE	102
Running time of electric energy meter	OPS_MONTHS	88
Asset model	MODEL_CODE	76
Manufacturing unit	MANUFACTURER	64

Table 3.2: Preserved Features

Feature Name	describe
A_1	The number of days the electric energy meter ran during the first abnormal collection
A_2	The region code of the payment terminal
A_3	Running time of electric energy meter
...	...
A_n	The wiring method of the electricity meter

We calculate the proportion of the importance of each feature in the total importance of all features as shown in equation 3.3.

$$P_{importance} = \frac{gini_j}{\sum_{i=1}^{n} gini_i} \tag{3.3}$$

In statistics, events with a probability of less than 4% are generally considered as low probability events. Here, 4% is selected as the proportion threshold for feature selection, and features with an importance greater than 4% are retained [18]. Model by retaining 12 features through filtering. All features are shown in Table 3.2, represented by symbol A.

Step three, construct features. The construction feature is based on business and expert experience, constructing new features for warning records and abnormal code records of electric energy meters according to business logic, and dividing the features into one vote veto feature and important feature [19].

Table 3.3: Construction features (partial)

Feature Name	describe
B_1	Is there any abnormality in the electricity meter
B_2	Does the abnormal setting of electricity meter rates occur
B_3	Does voltage exceeding the limit occur
...	...
B_n	Does the abnormality of phase B of the high supply and high meter occur

Table 3.4: Construction features (partial)

Feature Name	describe
C_1	Does the meter fly away
...	...
C_n	Does the electricity meter stop running
ALARM_CODE_0201	Does voltage phase failure occur

If there is an abnormality corresponding to a veto feature in an electric energy meter, then the meter must have malfunctioned; If there have been anomalies corresponding to important features in an energy meter, it is possible that the meter has malfunctioned. The construction features include 13 veto features and 30 important features. Some features are shown in Table 3.3, represented by symbol B.

Step 4 summarize a total of 55 feature fields mentioned above. This part of the features is shown in Table 3.4, represented by the symbol C, where $C = A \cup B$.

Based on the key links of the electricity meter, establish a fault rate prediction model according to the above characteristic data, and complete the batch fault prediction of the electricity meter. There are two solutions to predicting batch failure rates: One is to directly predict the failure rate of batch energy meters; The second is to predict whether a single meter has failed, and then calculate the failure rate of the batch of electricity meters based on the number and total number of failures in the batch.

The direct prediction of failure rate can be achieved by: (1) Using a regression model to fit the linear relationship between the features obtained and the batch failure rate. This method can obtain the optimal weight relationship between the batch failure rate and each feature; (2) By using the data from the split table to obtain failure rate data at different times, and applying a time series model, the trend prediction of batch failure rate on the timeline can be obtained.

Another approach is to first predict single table failures, and then divide the number of failures by the total number of batches to obtain the failure rate of the batch [20]. Predicting whether a single table is faulty is a binary classification problem, and simple classifier models such as decision trees, SVM, Bayesian, etc. can be used. The results of these models are easy to interpret, but their accuracy is average and they are prone to overfitting; Ensemble learning models can also be used, including random forest algorithm, XGBoost algorithm, lightgbm algorithm, etc, such models are integrated on the basis of simple models. Compared with a single model, they are often more accurate and can effectively avoid over fitting. However, the calculation rules are complex, and the interpretability of the model is poor.

Recording the process of random event changes and developments in chronological order constitutes a time series. Observing and studying the time series, searching for its patterns of change and development, and predicting its future trends is called Time Series Analysis.

Time series prediction only requires a set of historical data of the variables to be predicted. Compared with regression prediction models, this method does not require the effort to determine the causal relationship between variables, but only needs to extend the historical trend determined by the time series model outward to predict future changes. Time series prediction is often suitable for situations where the independent variable data required for regression models is relatively scarce, and the historical data of the variables to be predicted is relatively complete, which is sufficient to reflect their changing trends.

Regression analysis is a statistical analysis method aimed at determining the quantitative relationship of interdependence between two or more variables. According to the type of relationship between the independent and dependent variables, it can be divided into linear regression analysis and nonlinear regression analysis. In big data analysis, regression analysis is a predictive modeling technique that studies the relationship between the dependent variable (target) and the independent variable (predictor) [21]. This technique is commonly used for predictive analysis, time series modeling, and discovering causal relationships between variables.

Ensemble learning is a framework of machine learning that combines multiple models to improve their overall generalization ability. There are three types of ensemble learning: Bagging, Boosting, and Stacking. The XG Boost algorithm is an improved gradient boosting learning algorithm, which is a method in Boosting. The algorithm principle is different from the traditional GBDT algorithm. Traditional GBDT only utilizes first-order derivative information during the training process, while the XG Boost algorithm performs second-order Taylor expansion on the loss function and adds a regularization term outside the loss function to obtain the optimal solution. This not only ensures model accuracy but also limits the complexity of the model, avoiding overfitting. The XGBoost algorithm is based on a tree model. The XGBoost algorithm is an additive model composed of k base models. Assuming that the tree model we want to train in the t-th iteration is $f_t(x)$, the prediction result $widehaty_i^{(t)}$ of sample i in the t-th iteration satisfies equation 3.4. The model loss function satisfies the equation $\sum_{i=1}^{n} l(\widehat{y_i}, y_i)$. In the formula, n represents the number of samples.

$$\widehat{y_i^{(t)}} = \sum_{i=1}^{n} f_k(x_i) = \widehat{y_i^{(t-1)}} + f_t(x_i) \tag{3.4}$$

The definition of the model objective function is shown in equation 3.5. In the formula, t represents the number of trees, and Ω represents the regularization term.

$$Obj = \sum_{i=1}^{n} l(\widehat{y_i}, y_i) + \sum_{i=1}^{t} \Omega(f_i) \tag{3.5}$$

Performing a second-order Taylor expansion on equation 3.3 and removing the constant term yields the objective function for the t-th iteration as shown in equation 3.6. In the formula, g_i is the first derivative of the loss function, and h_i is the second derivative of the loss function.

$$Obj^{(t)} \backsimeq \sum_{i=1}^{n} [g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t) \tag{3.6}$$

According to equation 3.6, it can be seen that the XGBoost algorithm's loss function can be customized (there must be first and second derivatives), and the use of second derivatives makes gradient convergence faster and more accurate.

Establish linear regression, time series, and XGBoost algorithm models using existing Henan split table data. The hyperparameters of the three models are set to default values, and 98% is selected as the fault probability threshold. If the prediction result exceeds the threshold, it is judged as a fault Table, among them, linear regression and time series are used to determine batch fault tables, while XGBoost algorithm is only used for single Table fault determination. The model selected a total of 1190582 data from the first quarter of 2021 for training. Take archive information, R&D design data, material procurement data, production and manufacturing data, collected abnormal data, and measurement abnormal data as independent variables, and dismantle the sorting data to determine whether the electricity meter is faulty as the dependent variable to input into the XG Boost algorithm and linear regression model; Model batch failure rate as a time series. The model obtains the optimal joint probability distribution of the data through training, and applies this distribution to determine whether the smart energy meter is faulty in operation, achieving quality analysis of key links in the smart energy meter.

Under the same judgment criteria, the accuracy comparison of the three models is shown in Figure 3.2.

The XGBoost algorithm model has much higher prediction accuracy than other models, so the XGBoost algorithm model is chosen for subsequent analysis and prediction of electricity meter quality.

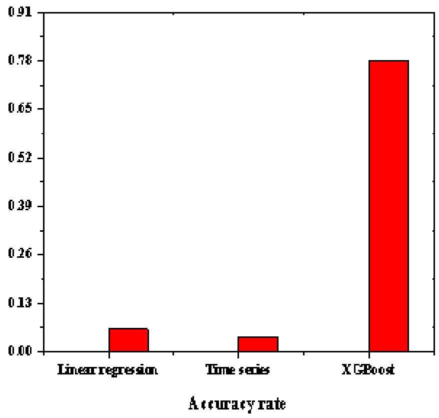

Fig. 3.2: Comparison of Model Accuracy

Table 4.1: Model Validation Results

Index	Value/piece	Proportion /%
Total	973621	
Actual number of faults	289178	8.00
Predict the correct quantity	695527	71.62
Number of prediction errors	276136	28.35
Predict the fault table as a fault Table(TP)	39342	4.06
Predict the fault table as a normal Table(FP)	48075	4.94
Predict a normal table as a faulty Table(FN)	61542	6.32
Predict the number of faults	100802	10.37

4. Result analysis. Using a total of 21157686 historical faulty electricity meter data from May 2019 to May 2021 in a certain area, a quality analysis model for intelligent electricity meters is established. The model predicts the fault data for three quarters from May to December 2021 in a certain area, and compares and verifies it with the actual dismantled data at the end of 2021.

The verification situation is as follows. The total number of electricity meters participating in the prediction is 3251317, involving 6112 arrival batches. Each meter is predicted for failure and compared with the actual results at the end of 2021 [22].

The validation data for the pre training stage of the model is that in the first quarter of 2021, there were 973621 electricity meters and 2313958 correctly predicted ones, accounting for 71.62%; The number of prediction errors is 917347, accounting for 28.35%. The detailed results are shown in Table 3.5.

After initial training, the accuracy of the model reached 0.73 and the recall rate reached 0.38. After verification, the model can recognize 44.65% of the fault tables, but there is also a 28.35% misjudgment situation. The accuracy of the model needs to be improved, and the evaluation indicators of the model performance are shown in Table 3.6 [23]. The definitions of model accuracy P and recall R are shown in equation 3.7:

$$P = TP/(TP + FP) \tag{4.1}$$

$$R = TP/(TP + FN) \tag{4.2}$$

After the first introduction of the model, the evaluation of the model results of training and test set, the accuracy of the training process is close to 1, while the accuracy of the test the index is stable around 0.70.

Table 4.2: Model Performance Evaluation Indicators

Evaluating indicator	Value
Accuracy	0.73
Accuracy	0.46
Recall	0.38

Finally, the validation process was evaluated and an accuracy of 0.46 was obtained. The results confirm that the general model is weak and overfitting may occur, requiring the optimization of the model parameters.

In order to improve the training model's performance, increase the generality, and reduce the risk of overwork, the model is gradually added to the electronic damage test data. set in the third and fourth quarters of 2021. pattern matching. Generalizability of the model to unknown data is verified using data augmented validation. Follow the pattern in two steps below:

1. The basic idea of good modeling based on K-fold cross-validation method is to group the original data in one way, one part is used as training set and the other factor is used as the test, and the classifier is first trained using the training method, and then as a performance test to evaluate the model to evaluate the training model using the test procedure. During the initial stage of model fitting, K-fold cross validation was used to improve the generalizability of the model [24].
   The main idea of K-fold cross-validation is to divide the original data equally into K sections, divide the data into K sections, select section i as the test set for section 3, and use the section K-1. based on the training set [25]. The average of the evaluation results of the K index was taken as the final evaluation of the model. The model is limited by this parameter in order to find the best combination for the model. Here, 5-fold cross-validation was used to select the correct one as the measurement parameter.
   With K-fold cross-validation, the precision of the training process was 0.66, the recall rate was 0.63, the measurement precision was 0.56, the recall rate was 0.55, and the acceptance precision was 0.46, and the reproducibility was increased. rate increased to 0.46. Up to 0.43, although the accuracy of training and testing has improved, the accuracy of the system is below 0.50, which is difficult to meet the needs, and it is necessary to take advantage of the model's hyperparameters;
2. Optimize the model a second time according to the network search method. Use the grid search method to optimize the hyperparameter values of the model.

The mesh search method involves partitioning the hyperparameters of the model into finite-valued elements. The program iterates over the composite values of all hyperparameters and selects the best model parameters based on parameter values as negative parameters.

Using the network search method, a precision of 0.82 for the training set, a recall rate of 0.84, a precision rate of 0.81, and a best return value of 0.78 for the parameters were obtained. 0.74 for the validation process and 0.82 for the recovery rate. A comparison of the effects before and after model development is shown in Figure 3.3.

The validation data of the model for each quarter from Q2 to Q4 2021 are shown in Figure 3.4.

Through optimization and adjustment, the model's parameters can quickly converge during the training phase, demonstrating excellent fitting ability in the training set. The accuracy rate in the validation set reaches 0.74, and the overall evaluation effect is relatively ideal, which can meet the actual needs of power grid business.

5. Conclusion. The author mainly focuses on the quality data of key links in electric energy meters, predicts the occurrence patterns of faults, and constructs an electric energy meter quality analysis model to study the quality analysis methods of key links in intelligent electric energy meters. The main research content includes the following two aspects:

1. Sort out the quality related data of key links in electric energy meters, study the key link data and quality impact feature extraction methods that affect the quality of electric energy meters, extract the regular features of faults in dismantled electric energy meters, use XG Boost algorithm model to learn the rules in dismantled electric energy meters, and construct a fault prediction model;

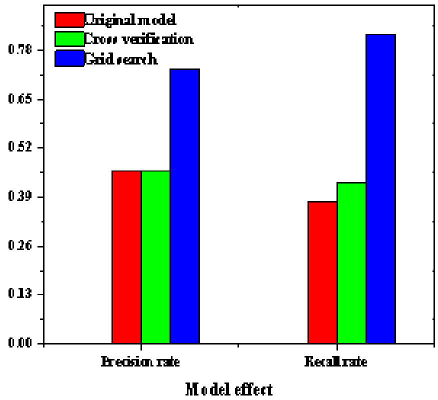

Fig. 4.1: Comparison of model effects before and after optimization

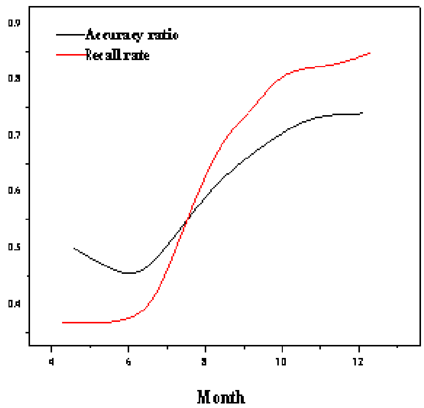

Fig. 4.2: Monthly Model Validation Results

2. Use the quality analysis model to identify the quality problem of the smart meter, model using the historical data, predict the explosion data at a specific location in April 2020, compare with the actual split data, and check. In May 2020, the model is optimized according to the confirmed results. The model was refined by using cross-validation and network research, and the final model achieved an accuracy of 0.74 and a recall rate of 0.82 of the validation process, which can be based on the real economy of the electricity project.

The author suggests a method to analyze and estimate the quality of smart meters during operation. This method is based on the important information from the main connection of the energy meter, and the selection of energy meter information for research and development, production, production, manufacturing, certification, installation, operation, demolition and disposal. sample data for design. The classification method of XGBoost algorithm is used to create intelligent models for the effective evaluation of power meters. After proving the accuracy, the results show that the accuracy of this method reaches 0.74, which can show the true value of the connection between smart meters.

Acknowledgement. The study was supported by Science and Technology Project funded by State Grid Co., LTD. (5700-202227226A-1-1-ZN).

REFERENCES

[1] Himeur, Y., Alsalemi, A., Al-Kababji, A., Bensaali, F., & Amira, A. (2020). Data fusion strategies for energy efficiency in buildings: Overview, challenges and novel orientations. Information Fusion, 64, 99-120.
[2] Ghazal, T. M. (2022). Energy demand forecasting using fused machine learning approaches. Intelligent Automation & Soft Computing, 31(1), 539-553.
[3] Kashinath, S. A., Mostafa, S. A., Mustapha, A., Mahdin, H., Lim, D., Mahmoud, M. A., ... & Yang, T. J. (2021). Review of data fusion methods for real-time and multi-sensor traffic flow analysis. IEEE Access, 9, 51258-51276.
[4] Muzammal, M., Talat, R., Sodhro, A. H., & Pirbhulal, S. (2020). A multi-sensor data fusion enabled ensemble approach for medical data from body sensor networks. Information Fusion, 53, 155-164.
[5] Sahu, A., Mao, Z., Wlazlo, P., Huang, H., Davis, K., Goulart, A., & Zonouz, S. (2021). Multi-source multi-domain data fusion for cyberattack detection in power systems. IEEE Access, 9, 119118-119138.
[6] Zhang, Y., Jiang, C., Yue, B., Wan, J., & Guizani, M. (2022). Information fusion for edge intelligence: A survey. Information Fusion, 81, 171-186.
[7] Jiang, L., Wang, X., Li, W., Wang, L., Yin, X., & Jia, L. (2021). Hybrid multitask multi-information fusion deep learning for household short-term load forecasting. IEEE Transactions on Smart Grid, 12(6), 5362-5372.
[8] Himeur, Y., Rimal, B., Tiwary, A., & Amira, A. (2022). Using artificial intelligence and data fusion for environmental monitoring: A review and future perspectives. Information Fusion, 86, 44-75.
[9] Chen, Y. (2021). College English teaching quality evaluation system based on information fusion and optimized RBF neural network decision algorithm. Journal of Sensors, 2021, 1-9.
[10] Nakutis, Ž., Rinaldi, S., Kuzas, P., & Lukočius, R. (2020). A method for noninvasive remote monitoring of energy meter error using power consumption profile. IEEE Transactions on Instrumentation and Measurement, 69(9), 6677-6685.
[11] Spertino, F., Chiodo, E., Ciocia, A., Malgaroli, G., & Ratclif, A. (2020). Maintenance activity, reliability, availability, and related energy losses in ten operating photovoltaic systems up to 1.8 MW. IEEE Transactions on Industry Applications, 57(1), 83-93.
[12] Karngala, A. K., & Singh, C. (2021). Reliability assessment framework for the distribution system including distributed energy resources. IEEE Transactions on Sustainable Energy, 12(3), 1539-1548.
[13] Zhou, M., Zhao, Y., Li, Q., Liu, X., Wang, Y., Sun, Y., ... & Yan, W. (2022). Smart Electricity Meter Reliable Life Prediction Based on Harmonic Factor. Electric Power Components and Systems, 50(13), 727-737.
[14] Ma, J., Teng, Z., Tang, Q., Qiu, W., Yang, Y., & Duan, J. (2021). Measurement error prediction of power metering equipment using improved local outlier factor and kernel support vector regression. IEEE Transactions on Industrial Electronics, 69(9), 9575-9585.
[15] Akhtar, I., Kirmani, S., & Jameel, M. (2021). Reliability assessment of power system considering the impact of renewable energy sources integration into grid with advanced intelligent strategies. IEEE Access, 9, 32485-32497.
[16] Zanghi, E., Brown Do Coutto Filho, M., & Stacchini de Souza, J. C. (2024). Collaborative smart energy metering system inspired by blockchain technology. International Journal of Innovation Science, 16(2), 227-243.
[17] Berka, A. L., MacArthur, J. L., & Gonnelli, C. (2020). Explaining inclusivity in energy transitions: Local and community energy in Aotearoa New Zealand. Environmental Innovation and Societal Transitions, 34, 165-182.
[18] Asselman, A., Khaldi, M., & Aammou, S. (2023). Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. Interactive Learning Environments, 31(6), 3360-3379.
[19] Chen, J., Zhao, F., Sun, Y., & Yin, Y. (2020). Improved XGBoost model based on genetic algorithm. International Journal of Computer Applications in Technology, 62(3), 240-245.
[20] Chakraborty, S., & Bhattacharya, S. (2021). Application of XGBoost algorithm as a predictive tool in a CNC turning process. Reports in Mechanical Engineering, 2(1), 190-201.
[21] Azmi, S. S., & Baliga, S. (2020). An overview of boosting decision tree algorithms utilizing AdaBoost and XGBoost boosting strategies. Int. Res. J. Eng. Technol, 7(5), 6867-6870.
[22] Ben Jabeur, S., Stef, N., & Carmona, P. (2023). Bankruptcy prediction using the XGBoost algorithm and variable importance feature engineering. Computational Economics, 61(2), 715-741.
[23] Song, P., & Liu, Y. (2020). An XGBoost algorithm for predicting purchasing behaviour on E-commerce platforms. Tehnički vjesnik, 27(5), 1467-1471.
[24] Li, S., & Zhang, X. (2020). Research on orthopedic auxiliary classification and prediction model based on XGBoost algorithm. Neural Computing and Applications, 32(7), 1971-1979.
[25] Sankar, S., Potti, A., Chandrika, G. N., & Ramasubbareddy, S. (2022). Thyroid disease prediction using XGBoost algorithms. Journal of Mobile Multimedia, 18(3), 1-18.

Edited by: Bradha Madhavan
Special issue on: High-performance Computing Algorithms for Material Sciences
Received: May 17, 2024
Accepted: Jun 10, 2024