

# Remote Malfunctional Smart Meter Detection in Edge Computing Environment

FANGXING LIU<sup>1</sup>, CHENGBIN LIANG<sup>2,3</sup>, AND QING HE<sup>1,2</sup>

<sup>1</sup>School of Instrumentation Science and Engineering, Harbin Institute of Technology, Harbin 150001, China

<sup>2</sup>National Institute of Metrology, Beijing 100001, China

<sup>3</sup>College of Electrical and Information Engineering, Hunan University, Changsha 410082, China

Corresponding author: Qing He (heqing@nim.ac.cn)

This work was supported by the National Key R&D Program of China under Grant 2016YFF0201201.

**ABSTRACT** Smart meter is a typical edge device that measures and records the energy data. The usage of smart meter data can improve the cyber physical relationship between the smart grid and cyber physical system. Millions of smart meters have been installed all around the world and the malfunction detection of large volume meter is a big issue. On site checking is a costing work and cannot meet the requirement for large scale meters. Online malfunctional meter detection and verification based on meter data analytics is a solution to the meter detection problem. For the purpose of detecting malfunctional smart meter, the low-voltage energy system model is studied and a meter error estimation method is proposed in this paper. This method adopts a decision tree to filter the abnormal data and classify data with different energy loss levels. Then clustering the data to obtain the data set with different energy usage behavior. A meter data matrix is constructed and meter error can be calculated from the solution of the matrix equation. A Recursive algorithm is adopted to solve the equation and estimate the meter error. The meter error above the regulation threshold will be classified as a malfunctional meter. The proposed approach has achieved higher accuracy in the experiment.

**INDEX TERMS** Smart meter, meter error estimation.

## I. INTRODUCTION

Smart grid integrates the power network infrastructure and cyber system and exhibits a typical nature of cyber physical system (CPS) [1]. Smart Grid is very complex and contains many subsystems. Many interconnected cyber assets associated with electric energy equipment and infrastructure are deployed in the energy grid. Smart grid aims to operate power technologies on the energy system to realize bidirectional power flows [2]. With the development of advance power technologies, the interconnected nature of sub-power system is becoming more complex and the cyber-physical component in energy grid is becoming more and more important.

Smart meters are important components in smart grid and the usage of smart meter data can improve the cyber physical relationship between smart grid and CPS. Smart meter can be used as sensors in the entire energy distribution grid because it can measure the energy usage and send the recorded data to energy management center [3]. The usage of smart meter

data can help to provide timely decisions for smart grid planning and operation through the CPS [4]. Smart meter data analytic can not only provide a mean to monitor the status of smart grid, but also can improve resilience against disruption and enhance grid asset management. Smart meter plays an important role in smart grid so that malfunctional smart meter detection will be meaningful work. The problem of abnormal metering performance caused by the scattered fixed installation of watt-hour smart meters remains exist. For a large number of smart meters and the complex application place, how to improve the ability to detect malfunctional smart meters has become the focus of attention of power grid enterprises [5], [6].

The development and establishment of Advanced Metering Infrastructure (AMI) in smart grid enables energy company to obtain a huge number of measurement data. Smart grid AMI provides a broad platform with its abundant measurement, communication, computing and storage resources and the analysis of those information has attracted much interest [7]. The solution for detecting malfunctional smart meter in a data analytic way are numerous. In [8], it presents a

The associate editor coordinating the review of this manuscript and approving it for publication was Xiaokang Wang.

generative model considers both hierarchical framework and meter data to detect abnormal smart meter. Fenza *et al.* [9] proposed an approach to define anomaly detection method ability in the face of concept drift. The approach adopts long short term memory network to profile and forecast consumer behavior. In [10], Cosmin described a comparison method consisting of the relationship between current and ideal power signature to detect anomaly devices. Zheng *et al.* [11] introduced a data mining technique to detect abnormal data and energy theft. The approach combined maximum information coefficient theory and clustering technique to find abnormal users with arbitrary shapes. In [12], Yang proposed a technique combine with Bollinger bands and partially observable Markov decision process to detect abnormal data and energy theft. A probabilistic adaptive model is adopted to improve efficiency. Zanetti *et al.* [13] presented a method where compares energy consumption before and after to detect fraud and abnormal usage. Xiao and Ai [14] presented an abnormal usage and energy theft detector based on random matrix theory. The correlation between energy consumption and abnormal system state is studied. In [15], it introduced an information fusion-based method to evaluate meter state online. The entropy weight technique is used to fuse sub-evaluation indicator to evaluate the state.

Meter error analysis is another way to detect malfunctional smart meter. Meter error is an import indicator of meter statuses. Two traditional approaches to detect abnormal meter by checking the meter error are on-site and laboratory checking. Energy company hires professional personnel take instruments and equipment to carry out a regular spot check. Also, the company can remove some meters from installation spot and take it to conduct professional testing in a laboratory. These two methods have a high detecting rate but with low management efficiency. It cost too much and cannot check all deployed smart meters. Estimating meter error online by processing meter reading data is a novel way to solve the large-scale checking problem. Korhonen [16] proposed a smart meter verification method by estimating meter error using consumption data. This method has less accuracy under large energy losses system. Kong *et al.* [17] introduced an online estimation method using a recursive least squares algorithm. A double-parameter method with dynamic forgetting factors was present to track meter parameter changes and calculate meter error. This estimation method has a high accuracy under meter data set without abnormal data. In [18], it proposed an algorithm that combines with cluster and regularization theory to estimate meter error. It functions well under low energy loss rate conditions. Kim *et al.* [19] proposed an intermediate monitor meter-based approach to detect fraud and abnormal data by solving linear system equations. The algorithm can also detect energy losses. In [20], a parameter degradation model was adopted to estimate meter error. The model had considered the effect factor that influences the meter error. In [21], Peng analyzed the distribution characteristics of meter error and influence quantity errors.

Monte Carlo approach was adopted to calculate the combined error.

In the related work, even though many methodologies have been proposed and some of the research work have achieved applications, there are still some limitations in real filed application. The idea of detecting abnormal meter using machine learning and reinforcement learning related algorithms is easy application but the many results depend on specific system structure and data set. Detecting malfunctional meter by checking meter error is an effective method. Online meter error estimation is a novel idea that can fulfill large-scale detection. But the error estimation accuracy is limited to data and system information. The more system feature and measurement data be collected, the higher estimation accuracy will be. However, some system information, such as energy loss, system power line resistance, are difficult to obtain in a real system. The absence of this information will influence the detection accuracy and method performance.

The goal of this paper is to detect malfunctional smart meter with fewer system information. This paper theoretically discusses the method of remote estimation of the measurement error of smart meter using sub-meter data and master meter data. The relationship between master meter and sub-meters based on AMI structure is analyzed. A data process framework combined with decision tree and cluster theory is proposed to filter abnormal data and gather similar feature data. Then a recursive least square algorithm is adopted to estimate meter error. The meter with large errors will be classified as malfunctional meter.

The two main contributions of our work are summarized below:

(1) We propose a model for detecting large scale smart meter error verification. Energy consumption data is used to construct a linear equation and meter error estimator can be derived from the solution of the equation. Hundreds of smart meters can be detected simultaneously. This model is applicable to the system where system information is incomplete. Energy loss-related data are not necessarily required that our model has higher applicability compare to the existing method.

(2) We introduce an algorithm combine with classification and recursion theory to calculate the meter error. Data with similar energy losses rate level is classified by decision tree and then cluster them to obtain the different profile data. Then construct the linear equation and solve it by recursive least square to obtain the meter error estimator.

The rest of this paper is structured as follows: Section II introduces the proposed model, data processing procedure and meter error estimation method. Section III presents the results and analysis of the experiments. Finally, we summarize our work in Section IV.

## II. METHOD

This section describes the framework architecture and detail models.

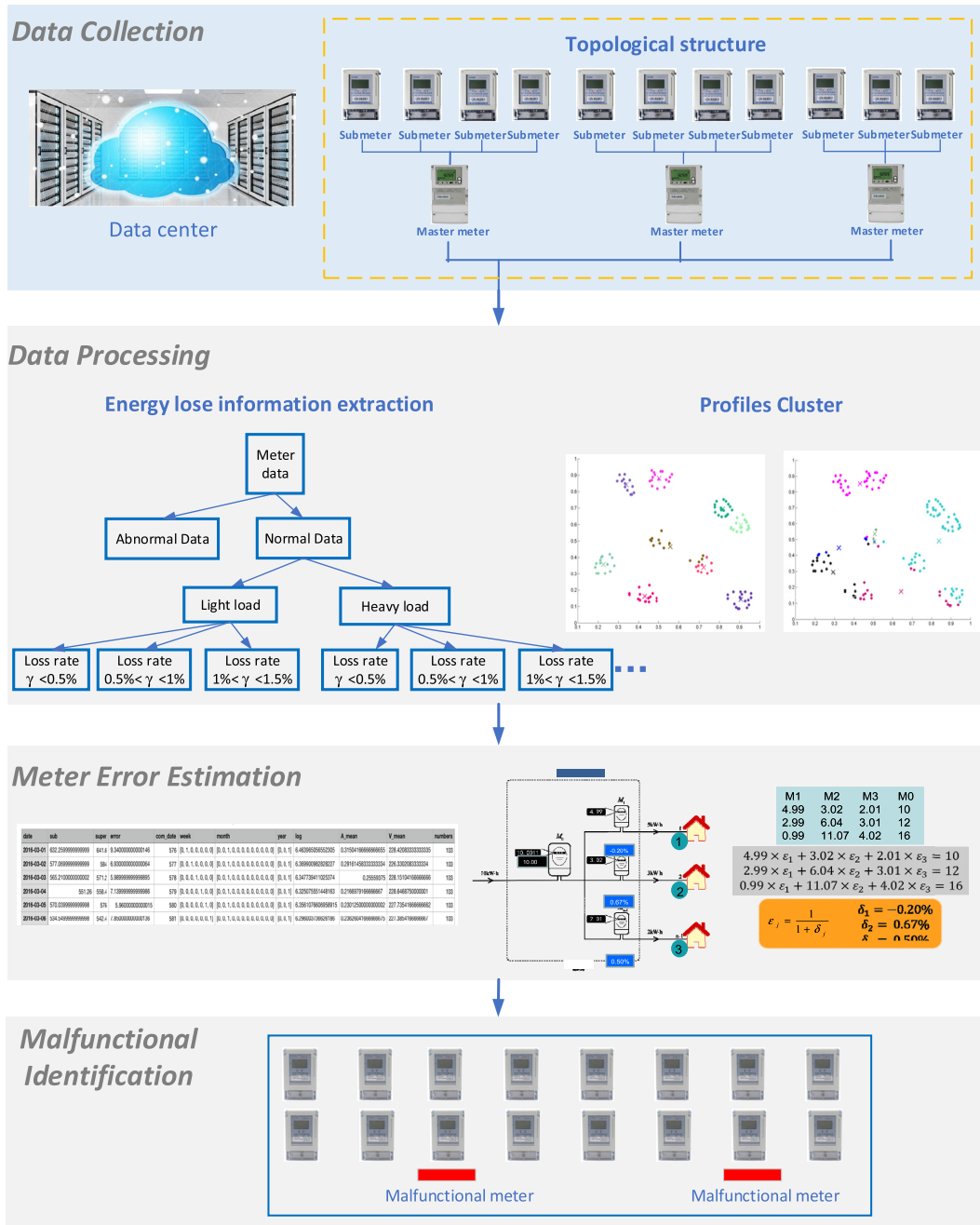


FIGURE 1. Workflow of meter detection.

A. METHOD OVERVIEW

The structure of the proposed method is shown in Figure 1. Energy data is recorded by smart meter and sent to the data center for storage. In a low voltage energy system, meters are deployed in a topological structure which means a high precision three-phase meter is installed in front of some low precision single-phase meters. The high-precision meter is the master meter and low-precision meter are submeters in the energy network. The submeter meter records voltage, current and energy consumption value of a residential user

while the total values of all users in an area are recorded by the master meter. The meter functions abnormally with inner components degradation and working environment changes. The meter reading error is an indicator widely used in the verification of smart meters by comparing the meter error with a threshold defined in official regulation. If meter error exceeds the threshold, the meter will be considered as a malfunctional meter. The aim of this our work is to detect the malfunctional meter by estimating meter reading error in an energy data analytics way. The meter error estimation

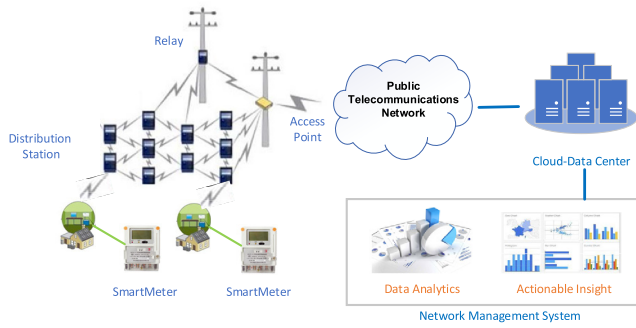


FIGURE 2. AMI data acquisition system.

approach includes four steps. Data collection is the foundational part of the work and the energy information should be acquired. The energy usage, voltage, current and system information such as meter ID, location number should be transmitted to the data center at a fixed frequency. Then processing data to remove the abnormal meter data and cluster representative profiles data. Next, constructing an equation according to the energy balance relationship and solve it to obtain meter error. Finally, the submeter with an estimated error above the regulation threshold is malfunctional meter.

**B. DATA COLLECTION**

The meter data is collected from AMI system. As shown in figure 2, the meter sends measurement data to access point over power line carrier, RS485 and wireless communication method. A concentrator usually installed at the access point to collect meter data and then upload to a cloud data center for further analysis. The data includes meter measurement data, such as energy consumption, voltage, current, and system information includes meter ID, meter number, date, time. Smart meter information has some specific characteristics. The official regulation specifies the threshold residual energy meter error is 2%.

**C. MODEL**

In the topological structure low voltage system, a master meter is deployed in front of submeters, as shown in Figure 1. Electric energy first passes through master meter then to submeters. In the same interval, the energy flow through master meter equals the sum of submeters in an ideal situation without any energy losses. This relationship can be written as:

$$\sum_{j=1}^n E_j(i) = E_0(i) \tag{1}$$

where  $E_j(i)$  is the true energy consumption on submeters breach  $i$  and  $E_0(i)$  is true total energy usage.

In a real energy system, the submeter has its measurement error because of inner components degeneration and some outside threats such as environment changes and abnormal installation. Energy loss exists in an energy system. It is also difficult to estimate it power loss. Considering energy loss

and meter error, then equation (1) can be written as:

$$\sum_{j=1}^n \frac{\phi_j(i)}{1 + \alpha_j} + E(i) = \phi_0(i) \tag{2}$$

where  $\alpha_j$  is the relative error of submeter  $j$ ,  $\phi_j(i)$  is the recorded energy consumption value of submeter  $j$ .  $\phi_0(i)$  is the usage value of master meter.  $E(i)$  is the energy loss in the system. The master meter has a higher accuracy than submeter error that its error could be neglected. The master meter value is considered as the true usage in the system. In  $n$  measurements, we can build a matrix equation:

$$\Phi^T \Theta = Y \tag{3}$$

where:

$$\Phi = \begin{bmatrix} \phi_1(1) & \phi_2(1) & \cdots & \phi_i(1) \\ \phi_1(2) & \phi_2(2) & \cdots & \phi_i(2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(n) & \phi_2(n) & \cdots & \phi_n(n) \end{bmatrix}$$

$$Y = \begin{bmatrix} \phi_0(1) - E(1) \\ \phi_0(2) - E(2) \\ \vdots \\ \phi_0(n) - E(n) \end{bmatrix} = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(n) \end{bmatrix} \Theta = \begin{bmatrix} \frac{1}{1 + \alpha_1} \\ \frac{1}{1 + \alpha_2} \\ \vdots \\ \frac{1}{1 + \alpha_n} \end{bmatrix} = \begin{bmatrix} \theta(1) \\ \theta(2) \\ \vdots \\ \theta(n) \end{bmatrix}$$

Meter error can be calculated from the solution of equation (3). Smart meter self-consumption  $e^M(i)$ , power line losses  $e^N(i)$  and leakage losses  $e^L(i)$ , are the main part of power losses. The mathematic formulation of energy losses  $E(i)$  is:

$$E(i) = e^M(i) + e^N(i) + e^L(i) + \varepsilon(i) \tag{4}$$

where  $\varepsilon(i)$  is the error term. The detail formulation of  $e^M(i)$ ,  $e^N(i)$ ,  $e^L(i)$  are:

$$e^M(i) = \sum_{j=0}^{n-1} \bar{p}_j t(i) \tag{5}$$

$$e^L(i) = U^2 \sigma t(i) \tag{6}$$

$$e^R(i) = \sum_{j=0}^{n-1} \bar{I}(i)_j^2 r_j t(i) \tag{7}$$

where  $p_j$  is the rated power of meter  $j$ ,  $\sigma$  is line leakage conductance.  $r_j$  is power line resistance from submeter  $j$  to master meter. In a real system,  $\sigma$  and  $r_j$  are difficult to obtain then losses are hard to estimate accurately. So  $e^N(i)$  and  $e^L(i)$  could be considered as part of the error term of energy losses in our model.

**D. DATA PROCESSING**

Data processing is an important work in our estimation model. In a real system, energy loss is difficult to estimate due to the lack of some parameters. As is shown in Equation (3), the energy loss term will have a significant impact

on the equation solution. If energy loss error term is greater than energy losses caused by all submeter errors, the error estimation will not inaccurate. The meter reading error of smart meter is not a constant value with the usage time grows. Meter error will change under different energy loads.

There are abnormal data in data sets. In the model master meter usage value should be larger than the sum of submeters. We should remove the data which sum of submeter reading excess the master meter for consideration of energy theft and power line break. To reduce the influence of energy loss and energy load on the solution of meter error, a method based on decision tree and clustering in proposed to process data.

The decision tree is widely used in classification research with smaller calculation workload and higher classification accuracy [22]. The data first divide into abnormal and normal data. Then classifying normal data into light load and heavy load. The light load data means the master meter current value is less than the meter rate current and heavy load is meter current greater than rate current. Then classifying data with different energy loss ratios. The ratio  $\gamma$  is:

$$\gamma = \frac{\phi_0(i) - \sum_{j=1}^n \phi_j(i)}{\phi_0(i)} \quad (8)$$

Data can be classified into different energy loss level with  $\gamma$  fall into different ratio range. The range is divided from  $(0, \gamma_{\max} / k]$  to  $((k-1) \gamma_{\max} / k, \gamma_{\max}]$ .  $k$  is the number of nodes.

Clustering is a classification method to aggregate data with similar features [23].  $K$  means clustering is a method for clustering centers in a set. Starting from choosing a desired cluster center number  $R$  and  $K$ -means procedure iteratively moves the centers to minimize the total distance between point and center.

For a given observation set  $x^{(i)} = \{x_1^{(i)}, x_2^{(i)}, \dots, x_n^{(i)}\}$  and cluster center with centroids  $c^{(j)} = \{c_1^{(j)}, c_2^{(j)}, \dots, c_n^{(j)}\}$ , the distance between observation ( $x^{(i)}$ ) with a centroid ( $c^{(j)}$ ) is:

$$dist_{cj}(x^{(i)}, c^{(j)}) = \|x^{(i)} - c^{(j)}\|_2 = \sqrt{\sum_{m=1}^n (x_m^{(i)} - c_m^{(j)})^2} \quad (9)$$

Given the cluster  $C = \{C^1, C^2, \dots, C^R\}$  whose centroids are  $c^{(j)}$ , cluster aim to minimize the distance of each point to the centroid.

$$\min \sum_{m=1}^n \sum_{x^{(i)} \in C^R} \|x^{(i)} - c^{(j)}\|_2 \quad (10)$$

The center is evaluated as:

$$c^{(i)} = \frac{\sum_{x^{(i)} \in C^R} x^{(i)}}{n_i} \quad (11)$$

where  $n_i$  is the number of observation points belonging to center  $c^{(i)}$ .

In our work, the decision tree method is used to identify the abnormal data preliminary and divide the normal data into

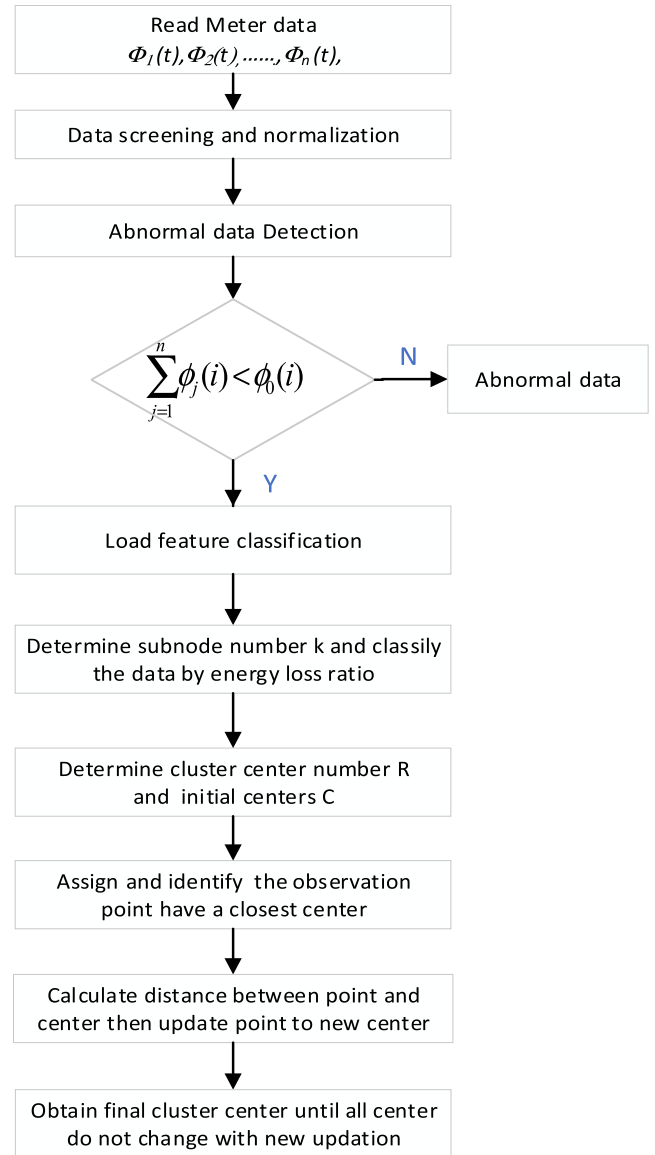


FIGURE 3. Flowchart of data processing.

a similar energy loss level part. Then clustering the data to obtain similar feature data for the meter error estimation.

Figure 3 shows the step of data processing.

### E. METER ERROR ESTIMATION

The submeter matrix is an  $N \times N$  dimensional matrix and the solution is a parameter related to meter error. The least square method is a common and effective method in the linear regression problem because of its simplicity and provides maximum likelihood regression estimators. The recursive model can update the parameter simultaneously and save storing resources. In order to update the estimated result repeatedly and improve the data usage efficiency, a recursive weight least squares model is used for solving the equation.

To minimize the residual sum of square error, the cost function of equation (3) is:

$$J = \sum_{i=1}^n (y(i) - \phi(i)\hat{\theta}(i))^2 \quad (12)$$

The recursive weight least squares solution for  $\theta$  is:

$$\hat{\Theta} = (\Psi^T(n)W(n)\Psi(n))^{-1}\Psi^T(n)W(n)Y(n) \quad (13)$$

where  $W = \text{diag}(w_1, \dots, w_n)$ .  $W$  is a weight factor matrix. The recursive  $\hat{\theta}(n + 1)$  is calculated as the sum of previous estimates. Then the recursive equation is:

$$\hat{\Theta}(n + 1) = \hat{\Theta}(n) + \zeta(n)(y(n + 1) - \Phi(n + 1)^T\hat{\Theta}(n)) \quad (14)$$

$\hat{\Theta}(n + 1)$  denotes the new estimator.  $\zeta(n)$  is the updated factor.

$$\hat{\Theta} = (\Psi^T(n)W(n)\Psi(n))^{-1}\Phi(n + 1) \quad (15)$$

Using exponential forgetting factor in weight matrix. Then the estimation equation is:

$$\hat{\Theta}(n + 1) = \hat{\Theta}(n) + P(n + 1)\Phi(n + 1)(y(n + 1) - \Phi(n + 1)^T\hat{\Theta}(n)) \quad (16)$$

$$\zeta(n) = P(n)\Phi(n + 1)(\Phi^T(n + 1)P(n)\Phi(n + 1) + \lambda)^{-1} \quad (17)$$

$$P(n + 1) = (I - \zeta(n)\Phi^T(n + 1))P(n)\frac{1}{\lambda} \quad (18)$$

where  $\lambda$  is the forgetting factor.

Root Mean Square Error (RMSE) is widely used to evaluate the estimation performance by calculating the distance between the estimator and real value. A smaller RMSE means a better performance. The definition is:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{\alpha}_i - \alpha_i)^2} \quad (19)$$

where  $\hat{\alpha}$  is the estimated error and  $\alpha_i$  is the true error value,  $n$  is the number of submeter.

### III. EXPERIMENT

The dataset is collected in a low voltage energy system from a residential community in urban. A master meter and 122 submeters are deployed in this energy system. The dataset is composed energy consumption value of all meters from August 2014 to August 2016 every 24 hours. Also, the master meter records the voltage and current value every 15 minutes. Most of the users are resident users in a community. The submeter has technical information: a rated power of 1100W, a rated voltage of 220V and a rated current of 5A. In addition to meter electricity measurement value, meter ID, system ID and community ID are recorded in our dataset. We removed empty and malformed rows, redundant data and invalid data in our data set, then extracted available data. Assume that all meter data are valid and no power theft and device failures in the system. All meter errors remain unchanged during this period. Then we randomly embedded malfunction meters

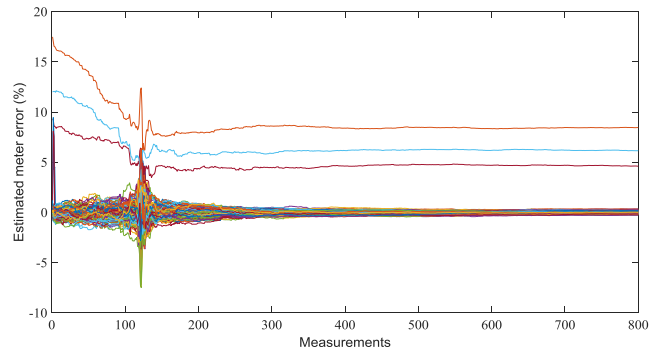


FIGURE 4. Estimated error in different measurements.

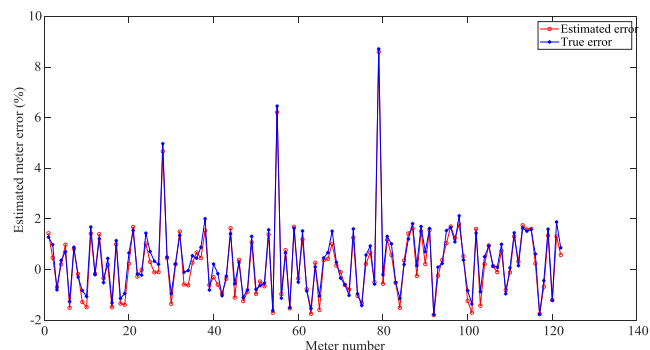


FIGURE 5. Meter error estimator.

by assigning large meter error which excess 2% to some submeters. The meter error estimator which greater than 2% will be considered as malfunction meter.

Begin with screening and classifying data, then clustering data and constructing the linear equation to estimate meter error.

#### A. RESULT OF METER ERROR ESTIMATION

The estimated meter error is shown in Figure 4. X-axis is the number of measurement and Y-axis represents the meter error. Meter error updated when new data input to model.

In Figure 4, each line shows the error estimator of every recursion. The estimator converges to stable value as the measurement increases. Three meters have significantly high error than other meters and some estimated error very close to 0. The number of measurements required depends on the data characteristics. It takes around 300 measurements to find a stable value for error estimator. The recursive model works well to estimate meter error with little variation.

Figure 5 shows the detail information of error estimator when all values remain stable. It can be seen that meter 28, meter 55 and meter 79 have large meter errors that exceed 2%. Meter 79 had the largest meter error 8.596%, while meter 55 and meter 28 had an error estimator of 6.213% and 4.669% respectively. The other meters have an error estimator fall into a normal range from -2% to 2%. The RMSE of this method is 0.21% which means the estimated error has a very small

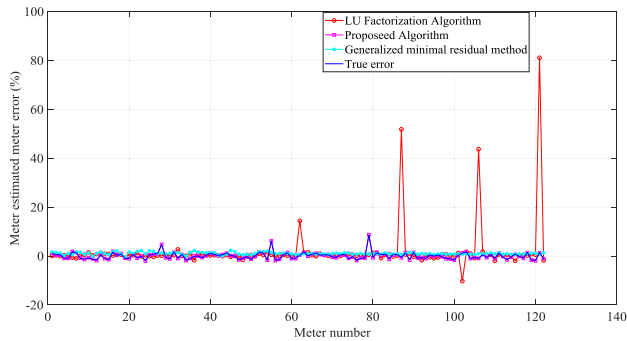


FIGURE 6. Meter error estimator comparison.

TABLE 1. Performance comparison.

	LU factorization	GMRES	Our method
RMSE	92%	0.9%	0.21%

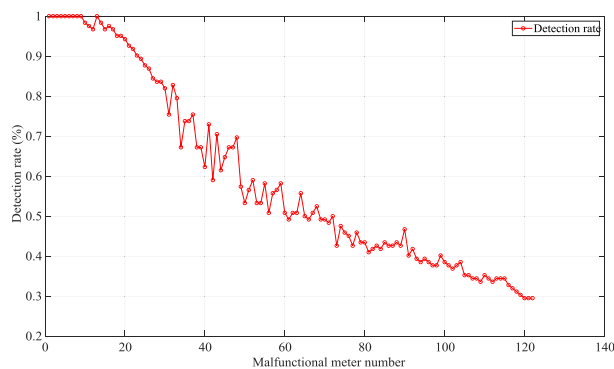


FIGURE 7. Detection rate.

deviation from the true error value. The error estimator can express the error characters of meter.

### B. COMPARISON WITH CLASSICAL METHODS

We compared our approach with LU factorization and generalized minimal residual method (GMRES) with same data. Figure 6 shows that the LU factorization has a non-convergence estimator due to the ill-condition of sub-meter matrix. The error estimator obtained from generalized minimal residual method contains the singular value.

Table 1 shows the *RMSE* of the three methods. The proposed method has the lowest value while LU factorization has the largest *RMSE*. The meter error threshold is 2%. The *RMSE* should not be greater than 2%. LU factorization approach shows bad performance in our dataset and our model performs better than GMRES due to a smaller *RMSE* value.

### C. DETECTION RATE ANALYSIS

Figure 7 demonstrates the detection rate of our model according to the different numbers of malfunctional meters in the system. The detection rate refers to the number of meter status

correctly classified to the total meter number. It can be seen that our model works well when malfunctional meter number is not greater than 23. The detection rate is above 90%. All abnormal data can be detected if no more than 10 malfunctional deployed in the system. The model performance is getting worse with the increase of malfunctional meter.

## IV. CONCLUSION

This paper proposes a data analytics approach to detect the malfunctional smart meter. This method requires huge data and can be applied in an edge computing environment. In this paper, a low voltage smart meter system has been explained and a recursive algorithm based on decision tree and cluster theory is proposed to estimate meter error. The experiment results show the excellent performance of our approach. Smart meter is a typical edge device. Smart meter data analytics on meter detection has widely application and huge economic value. We hope this paper can provide readers a picture meter error online estimation and smart meter data analytics in edge computing environment.

## REFERENCES

- [1] X. Yu and Y. Xue, "Smart grids: A cyber-physical systems perspective," *Proc. IEEE*, vol. 104, no. 5, pp. 1058–1070, May 2016.
- [2] J. E. Dagle, "Cyber-physical system security of smart grids," in *Proc. IEEE PES Innov. Smart Grid Technol. (ISGT)*, Jan. 2012, pp. 1–2.
- [3] J. Ekanayake, N. Jenkins, K. Liyanage, J. Wu, and A. Yokoyama, *Smart Grid: Technology and Applications*. Hoboken, NJ, USA: Wiley, 2012.
- [4] C. A. Macana, N. Quijano, and E. Mojica-Nava, "A survey on cyber physical energy systems and their applications on smart grids," in *Proc. IEEE PES Conf. Innov. Smart Grid Technol. LATIN Amer. (ISGT LA)*, Oct. 2011, pp. 1–7.
- [5] Q. Sun, H. Li, Z. Ma, C. Wang, J. Campillo, Q. Zhang, F. Wallin, and J. Guo, "A comprehensive review of smart energy meters in intelligent energy networks," *IEEE Internet Things J.*, vol. 3, no. 4, pp. 464–479, Aug. 2016.
- [6] D. Alahakoon and X. Yu, "Smart electricity meter data intelligence for future energy systems: A survey," *IEEE Trans. Ind. Informat.*, vol. 12, no. 1, pp. 425–436, Feb. 2016.
- [7] N. Saputro and K. Akkaya, "Investigation of smart meter data reporting strategies for optimized performance in smart grid AMI networks," *IEEE Internet Things J.*, vol. 4, no. 4, pp. 894–904, Aug. 2017.
- [8] R. Moghaddass and J. Wang, "A hierarchical framework for smart grid anomaly detection using large-scale smart meter data," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 5820–5830, Nov. 2018.
- [9] G. Fenza, M. Gallo, and V. Loia, "Drift-aware methodology for anomaly detection in smart grid," *IEEE Access*, vol. 7, pp. 9645–9657, 2019.
- [10] C. Cernazanu-Glavan and M. Marcu, "Anomaly detection using power signature of consumer electrical devices," *Adv. Electr. Comput. Eng.*, vol. 15, no. 1, pp. 89–94, 2015.
- [11] K. Zheng, Q. Chen, Y. Wang, C. Kang, and Q. Xia, "A novel combined data-driven approach for electricity theft detection," *IEEE Trans. Ind. Informat.*, vol. 15, no. 3, pp. 1809–1819, Mar. 2019.
- [12] Y. Liu and S. Hu, "Cyberthreat analysis and detection for energy theft in social networking of smart homes," *IEEE Trans. Comput. Social Syst.*, vol. 2, no. 4, pp. 148–158, Dec. 2015.
- [13] M. Zanetti, E. Jamhour, M. Pellenz, M. Penna, V. Zambenedetti, and I. Chueiri, "A tunable fraud detection system for advanced metering infrastructure using short-lived patterns," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 830–840, Jan. 2019.
- [14] F. Xiao and Q. Ai, "Electricity theft detection in smart grid using random matrix theory," *IET Gener., Transmiss. Distrib.*, vol. 12, no. 2, pp. 371–378, Jan. 2018.
- [15] H. Cai, H. Chen, X. Ye, X. Zhang, H. Wen, J. Li, and Q. Guo, "An online state evaluation method of smart meters based on information fusion," *IEEE Access*, vol. 7, pp. 163665–163676, 2019.

[16] A. Korhonen, "Verification of energy meters using automatic meter reading data," in *Proc. Named Entities Workshop*. Stroudsburg, PA, USA: Assoc. Comput. Linguistics, 2012, pp. 1–2.

[17] X. Kong, Y. Ma, X. Zhao, Y. Li, and Y. Teng, "A recursive least squares method with double-parameter for online estimation of electric meter errors," *Energies*, vol. 12, no. 5, p. 805, 2019.

[18] L. Fangxing, H. Qing, H. Shiyang, W. Lei, and J. Zhengsen, "Estimation of smart meters errors using meter reading data," in *Proc. Conf. Precis. Electromagn. Meas. (CPEM)*, Paris, France, Jul. 2018, pp. 8–13.

[19] J. Y. Kim, Y. M. Hwang, Y. G. Sun, I. Sim, D. I. Kim, and X. Wang, "Detection for non-technical loss by smart energy theft with intermediate monitor meter in smart grid," *IEEE Access*, vol. 7, pp. 129043–129053, 2019.

[20] Z. Zhang, H. Gong, C. Li, Z. Wang, and X. Xiang, "Research on estimating method for the smart electric energy meter's error based on parameter degradation model," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 366, Jun. 2018, Art. no. 012065.

[21] P. Xiaojuan, W. Lei, J. Zhengsen, W. Xuwei, H. Hongtao, and L. Lijuan, "Evaluation of digital energy meter error by Monte Carlo method," in *Proc. Conf. Precis. Electromagn. Meas. (CPEM)*, Jul. 2016, pp. 1–2.

[22] E. Hullermeier, "Experience-based decision making: A satisficing decision tree approach," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 35, no. 5, pp. 641–653, Sep. 2005.

[23] S. Na, L. Xumin, and G. Yong, "Research on k-means clustering algorithm: An improved K-means clustering algorithm," in *Proc. 3rd Int. Symp. Intell. Inf. Technol. Secur. Informat.*, Apr. 2010, pp. 63–67.



**CHENGBIN LIANG** was born in Hebei, China, in 1994. He received the M.Sc. degree in mathematics from Guizhou University, Guiyang, China, in 2018. He is currently pursuing the Ph.D. degree in electrical engineering with Hunan University, Changsha, China. His research interests include power quality analysis and data mining for smart meters.



**FANGXING LIU** received the B.S. and M.S. degrees from the Harbin Institute of Technology, where he is currently pursuing the Ph.D. degree with the School of Instrumentation Science and Engineering. His main research interests include smart meter data analytic and meter calibration.



**QING HE** was born in Beijing, China, in 1965. He received the B.Sc. degree from Tsinghua University, Beijing, in 1988, the M.Sc. degree from the Beijing University of Chemical Technology, Beijing, in 1999, and the Ph.D. degree from the Harbin Institute of Technology, Harbin, China, in 2009. He is a part-time Professor with the Harbin Institute of Technology, Harbin. He joined the National Institute of Metrology, Beijing, in 1988, where he was involved in the establishment of Chinese quantum hall system standard. His current research interests include the joule balance, ac quantum Hall resistance, quantum device fabrication, and new electromagnetic measurement standards for China under big data.

• • •