World Wide Web (2020) 23:1177–1195
https://doi.org/10.1007/s11280-019-00705-w

Intelligent maintenance frameworks of large-scale grid using genetic algorithm and K-Mediods clustering methods

Weifeng Wang[1] · Bing Lou[2] · Xiong Li[1] · Xizhong Lou[3] · Ning Jin[3] · Ke Yan[3]

Received: 22 January 2019 / Revised: 5 May 2019 / Accepted: 1 July 2019 /
Published online: 13 July 2019

Abstract

Large-scale power grids, especially smart grid systems, consist of a huge amount of complex computerized electronic devices, such as smart meters. A smart maintenance system is desired to schedule and send maintenance worker to locations where any computerized devices become faulty. A grid management system (GMS) is purposely designed in the way that the following three conditions are generally fulfilled: 1) all workers are working on full capacity everyday; 2) the highest severity level faulty devices are fixed the quickest; and 3) the overall traveling distance/time is minimized. In this study, two intelligent grid maintenance framework are proposed considering the above three conditioned based on two state-of-arts algorithms, namely, genetic algorithm and K-mediods clustering method, respectively. Five real-world datasets collected from five different local cities/counties in eastern China are adopted and applied to verify the effectiveness of the two proposed intelligent grid maintenance frameworks.

Keywords Smart electric power grid · Maintenance planning · Genetic algorithm · K-mediods clustering

1 Introduction

Electric power and its grid development are among the most important living bases for people living and civilizations of the modern world. More recently, the development of next

This article belongs to the Topical Collection: *Special Issue on Smart Computing and Cyber Technology for Cyberization*
Guest Editors: Xiaokang Zhou, Flavia C. Delicato, Kevin Wang, and Runhe Huang

✉ Ke Yan
yanke@cjlu.edu.cn

[1] State Grid Zhejiang Electric Power Co., Ltd, Hangzhou, 310007, China

[2] Zhejiang Huayun Information Technology Co., Ltd, Hangzhou, 310008, China

[3] College of Information Engineering, China Jiliang University, 258 Xueyuan Street, Hangzhou, 310018, China

generation power distribution system, i.e., the smart grid, has attracted enormous attentions across the world. The concept of smart grid integrates the power of traditional grid technology with cutting-edge computer and internet technology to build a intelligent power system that receives, analyzes and applies data from various sources [5, 23].

A typical smart grid system usually consists of a huge number of equipment devices, such as smart meters, smart appliances, renewable energy resources and energy efficient control devices [8]. Due to the large number of smart devices and the complexity of the device structure, failures occur frequently, which becomes a challenge of smart grid development for real-world applications, especially for developing countries with large geographic areas and populations [9, 14]. A grid maintenance system (GMS), is desired to identify fault locations, evaluate conditions and make decisions [18]. It is a standard procedure to send out workers and fix the faulty devices manually whenever necessary. However, due to the huge amount of the faulty devices, an intelligent maintenance planning is always demanded to reduce the time, manpower and economic losses.

This study intends to provide feasible solutions to the problem of maintaining a large-scale power grid in eastern China with limited number of maintenance workers and budget for a local company, named State Grid Zhejiang Electric Power LTD (SGZEP). The current grid maintenance system generally considers the following three criteria in maintaining the local power grid in China:

1. **Maximizing the working capacity for each working crew.** Adequate number of faulty devices that assemble in a certain area of region has to be assigned to each working crew to maximize his/her working capacity. The intelligent planning for each working crew maximizes the overall working capacity, and consequently reduces the necessary number of employing new working crews.
2. **Making the highest severity level faulty devices reparation at the top priority.** Each faulty device is marked with a severity rate/level that indicates the fault severity of that particular device. The top priority is always given to the higher severity level fault devices for each working crew when a region or an assembly of devices is assigned to him/her.
3. **Minimizing the traveling distance/time for each working crew.** For each working package that is assigned to a particular working crew, there is always a minimal distance traversal path realizable for the working crew, such that, on that path, every neighboring pair of the faulty devices is close enough to each other. The overall traveling distance or time must be minimized from the GMS point of view.

In this work, two intelligent grid maintenance framework are proposed considering the above three factors based on two state-of-arts algorithms, namely, genetic algorithm (GA) and K-mediods clustering method, respectively. The GA based maintenance framework (GAMF) considers each working crew at one time and searches for the optimal working path that maximizes the overall severity level and simultaneously minimizes the overall traveling time. An overall management planning scheme is proposed to iteratively assign workers one at a time. The K-medoids clustering based maintenance framework (KMCMF) clusters all faulty devices into k clusters, where k is the number of available working crews on each day. The customized algorithm is carefully designed in the way that each cluster contains one top k severity level device; and all faulty devices are reasonably close to each other within each cluster. After all faulty devices are clustered, each working crew will be assigned with one cluster and works starting from the most sever faulty device in that cluster. The criterion 2 is considered first followed by criteria 3 and 1.

A real-world dataset collected by the local electric power gird maintenance company in eastern China, including maintenance scenarios in several local cities in China, is utilized to verify the effectiveness of the proposed algorithms in the experiment section. The power grid comprises approximately 25 million smart power meters; and perform computerized analysis as well as Web services based on data sent back from the smart meters. However, almost on each day, a number of devices becomes faulty because of various reasons. In tradition, the local power grid company sends out reparation working crews on each day based on trivial strategies, such as the shortest distance path first algorithm or highest severity level first algorithm. In this study, we compare our two proposed maintenance working strategies with the trivial algorithms to show the advantages of using intelligence enhanced strategies in maintaining large-scale power grid systems.

1.1 Problem specification

Considering a large-scale power grid system covering a medium size city consisting of approximately 25 million smart meters, on each day, there are in total n faulty meters. All faulty meters are enumerated as m_1, m_2, ..., m_n, where m_i is an entity containing the information about the device location, fault type, severity level, approximated fixing hours and etc. On each day, there are k available reparation working crews. All working crews are enumerated as p_1, p_2, ..., p_k, where p_j is an entity containing the worker's personal information, traveling speed and location. The problem arises when the number the faulty devices always greatly exceeds the number of available working crews, i.e., $k << n$. An intelligent maintenance scheme of assigning faulty smart meters to particular working crew is desired considering the three criteria, namely, the overall capacity, overall severity level and overall traveling distance. Currently, the local company alternatively chooses between two traditional approaches to send working crews to fix faulty devices everyday, which are shortest distance path first strategy or highest severity level first strategy. A more sophisticated strategy is demanded to consider the three criteria concurrently.

1.2 Contributions

The main contributions of this work include:

- **Discretizing severity levels for different faulty devices.** Discretization of severity levels of faulty devices is an important step for intelligent control, maintenance, fault diagnosis and other related industrial applications. In this study, we introduce a realistic way of calculating the severity level from the grid company's point of view. The severity level calculation involves three factors, including the fault lasting hours, averaged power consumption each day for the faulty meter and the number of days towards the next billing day. Compared with our previous work in [39], the calculation formula is formal; and all severity levels are not bounded.
- **Proposing two intelligent maintenance algorithms.** Two intelligent maintenance algorithms are proposed to automatically assign faulty devices to particular working crews, namely, genetic algorithm based maintenance framework (GAMF) and K-medoids clustering based maintenance framework (KMCMF). A series of experiments are conducted to verify the effectiveness of the proposed two algorithms.
- **Applying theoretical artificial intelligence (AI) algorithms to a real-world problems.** Both genetic algorithm and K-medoids clustering algorithm are AI enhanced

algorithms that have been widely applied to many scientific research areas, such as bioinformatics, image processing, human computer interaction and etc. However, there are quite few chances that we can verify the theoretical algorithms in real-world industrial applications. In this work, five real-world datasets collected from five different local cities/counties in eastern China are adopted and applied to verify the effectiveness of the two proposed algorithms.

2 Related work

Fault detection, diagnosis, evaluation and reparation (FDDER) is an important research topic in both scientific and industrial areas [16, 17, 28, 31]. As early as 1988, Viswanadham and Johnson [38] started to develop intelligent monitoring and control system to perform automatic fault detection and diagnosis for manufacturing systems. In 1990, Nelson [28] introduced the concept of fault tolerance, which emphasized the importance of fault evaluation in FDDER. All faults were then classified into different severity levels. For low severity levels, immediate reparation is not required; and fault tolerance is used. Reparation working crews have the choice to work on high severity level faulty devices based on the system evaluation. In 2006, Jardine *et al.* reviewed the state-of-art technologies in the field of condition-based maintenance. A maintenance recommendation is made by the management system usually based on three steps: data acquisition, data processing and maintenance decision-making.

More recently, due to the rapid development of Internet, database and artificial intelligence (AI) technology, the data size is growing exponentially. The concept of big data is introduced, along with the AI technology development in areas, such as data mining, machine learning and deep learning [1, 7, 10, 11, 19, 27, 41, 42]. In 2006, Lee *et al.* [22] introduced the existing intelligent prognostics tools for e-maintenance of industrial subsystems. Qin [29] surveyed the more recent developments of data-driven methods for FDDER. Cai *et al.* [6] applied Bayesian networks to diagnose various faults in engineering systems. Yan *et al.* [44, 45] employed various machine learning techniques to diagnose fault of air handling unit subsystems in large-scale air-conditioning systems.

Genetic algorithm (GA) and K-medoids clustering method are two important technique in the field of machine learning and AI. GA and its extensions have been widely used in medical diagnosis, image processing and human computer interaction [3, 15, 25]. Global optimal solutions can be found by GA with multiple criteria formulated using fitness functions [26, 40]. Fei and Zhang [12] utilize genetic algorithm to diagnose potential faults inside power transformers. Samanta *et al.* performed experiments to detect bearing faults using two GA-based hybrid algorithms. One of them combines GA with artificial neural networks (ANNs); and the other combines GA with support vector machine (SVM). Samuel and Rajan proposed a hybrid particle swarm optimization (PSO) based genetic algorithm and a hybrid PSO based shuffled frog leaping algorithm to maintain a power system in long-term power generation scenarios. The proposed methods have been proved to be effective using real-world datasets.

K-medoids clustering algorithm is one of the most commonly used clustering algorithms, which has been widely applied to various industrial applications [13, 24, 37]. Rai and Upadhyay [30] utilized K-medoids clustering algorithm to evaluate bearing performance degradations. Bishnu and Bhattacherjee [4] applied K-medoids clustering algorithm to software fault prediction. For problems without absolute labeling system applied, unsupervised learning algorithms, such as the K-medoids clustering, can be helpful for sub-optimal predictions. For example, Zhong et al. [47] introduced to use an unsupervised learning algorithm, namely, generative adversarial neural network (GAN), to detect, maintain and diagnose possible faults in air handling units.

3 Severity level calculation for smart faulty meters

Under the situation where a large quantity of faulty smart meters are available, the maintenance management system is supposed to evaluate the severity level for each faulty meter before the faulty meters being assigned to a specific working crew. In this section, a discrete formulation of the severity level measurement is proposed for each faulty smart meter based on real-world data. Three important factors are considered in the formulation:

- The averaged power consumption of that particular unit on each day (k);
- The fault lasting hours (t);
- The number of days towards the next billing day (d).

3.1 Factor of averaged power consumption

The power consumption data collected by SGZEP consists of over 25 million of power consumption units, including households, small/medium companies and large-scale industrial factories. For normal households, the average power consumption in each month is likely below 200 kwh. For small/medium companies, the power consumption in each month is mostly in between of 200 and 1000 kwh; and for large-scale industrial companies, the power consumption can be above 1000 kwh. In this study, we averaged all units' monthly power consumption data in 2018, and define the factor of averaged power consumption for each unit as:

$$r(x_i) = \begin{cases} 1, & \text{for } g(x_i) < 200; \\ 2, & \text{for } 200 < g(x_i) < 1000; \\ 3, & \text{for } 1000 < g(x_i). \end{cases} \tag{1}$$

In this study, 10,000 residential units and 10,000 non-residential units are randomly selected with their monthly averaged power consumption in 2018, which is denoted as $g(x_i)$. We plot the graphs for these 20,000 units' power consumption deviation of the month May, 2018 in Figures 1 and 2. For residential units, 92% of the selected 10,000 units has power consumption deviation less than 17. And for the non-residential units, 88.75% of the selected 10,000 units has power consumption deviation less than 17. All these facts suggest that the value of $g(x_i)$ does not change that much over months. The formulation of $g(x_i)$ is proposed as:

$$g(x_i) = \alpha \cdot g(x_{i-12}) + (1 - \alpha) \cdot g(x_{i-1}), \tag{2}$$

where α is a parameter that can be tuned for each particular unit. The current value of $g(x_i)$ is assumed to be closely related to the power consumption in the same month, last year $g(x_{i-12})$ and in the last month, same year $g(x_{i-1})$.

3.2 Factor of fault lasting hours

The reparation of a faulty meter cannot be immediate; and the working crews are demanded to work on the most sever faulty device that was pre-scheduled by a management system. The time interval between the device becoming faulty and been fixed is another important factor to measure the severity level of a faulty device. Suppose that the meter reading of a

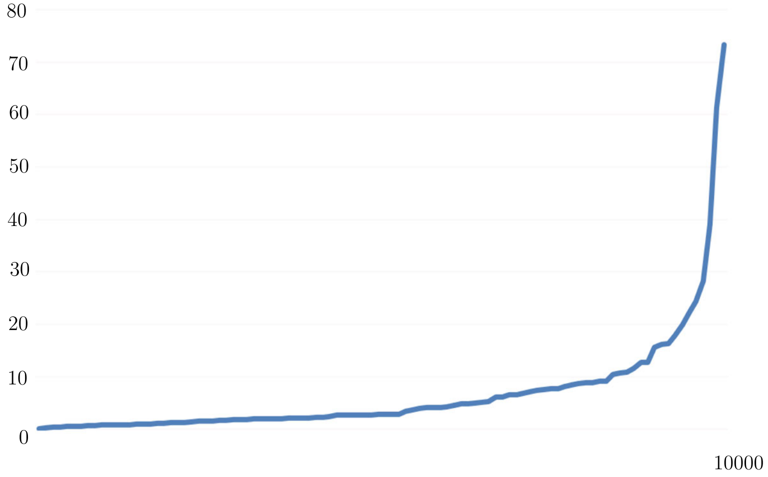


Figure 1 The deviations of randomly selected 10,000 residential units from $g(x_i)$ to the power consumption in May 2018. All deviations are sorted following ascending order

faulty meter is always invalid. Further assumption is made that the faulty reading cannot be used in any analysis and must be abandoned. The lasting hours of that particular fault determines the number of missing entries in the data management system. And the overall amount of power consumption that is missing can be proportional to the severity level:

$$j(x_i) = \frac{n}{\theta_i} \cdot r(x_i), \tag{3}$$

where θ_i is the number of days in the current billing month.

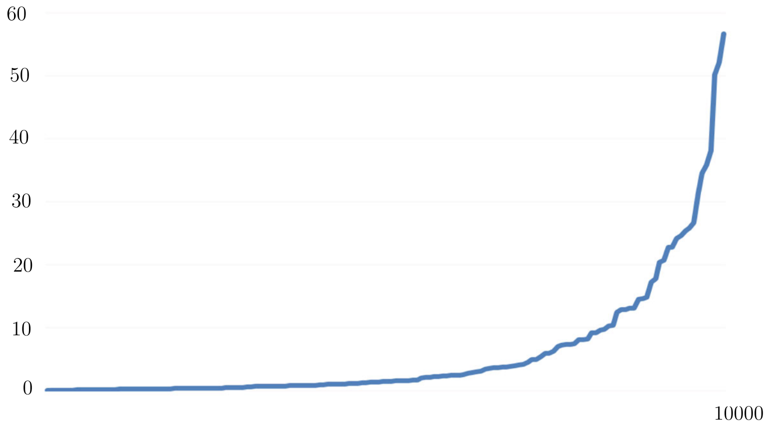


Figure 2 The deviations of randomly selected 10,000 non-residential units from $g(x_i)$ to the power consumption in May 2018. All deviations are sorted following ascending order

3.3 Factor of days toward next billing day

In most of the less developed countries, billing day is critical for a grid company to survive in a highly competitive environment. If the faulty meter is not fixed on the billing day, the exact billing amount cannot be ensured for the particular unit. A series of consequences might happen when the estimated billing amount is not matching the actual bill. Therefore, most of the faulty meters have to be fixed before the next billing day. In other words, the number of days towards the next billing day is another important factor for the severity level measurement.

To ensure the proposed method to be practical, we analyzed the number of faulty meters at 0:00, 1st of June, 2018 in the dataset collected by SGZEP. Figure 3 shows the number of units with consecutive missing metering data. Over 50,000 units with faulty meters have missing metering data for two days; and only around 10,000 units have missing data more than 15 days, which suggests that most of the faulty meters can be fixed within a short period of time, e.g., a week. In fact, according to the statistics, 59.06% of the faulty meters can be fixed within 8 days; the remaining 30% of the faulty meters can be fixed from 9 to 21 days; and only 10.22% of the faulty meters are fixed more than 22 days. For those faulty meters whose next billing day is more than 8 days away from the current date, the contribution to severity level can be small. Otherwise, the contribution to severity level increases when the next billing date is approaching. The last factor of the severity level measurement is defined as:

$$s(x_i) = \frac{9-l}{\theta_i} \cdot r(x_i), \tag{4}$$

where l is the number of days away from the next billing day.

3.4 Discretized severity level formulation

In summary, in this section, we proposed a discretized severity level formula for intelligent maintenance of a smart grid. The severity level becomes another important factor

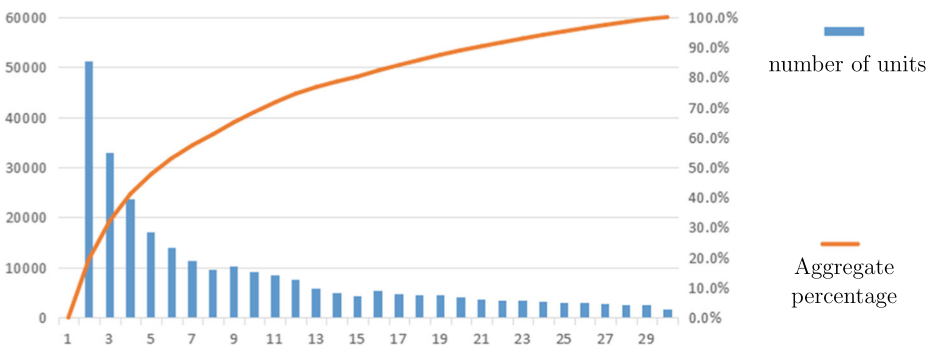


Figure 3 The number of faulty meters at 0:00, 1st of June, 2018 in the dataset collected by SGZEP. We sub-categorized all numbers by the fault lasting days. Around 60% of the faulty meters can be fixed within 8 days; the remaining 30% of the faulty meters can be fixed from 9 to 21 days; and only 10% of the faulty meters are fixed more than 22 days

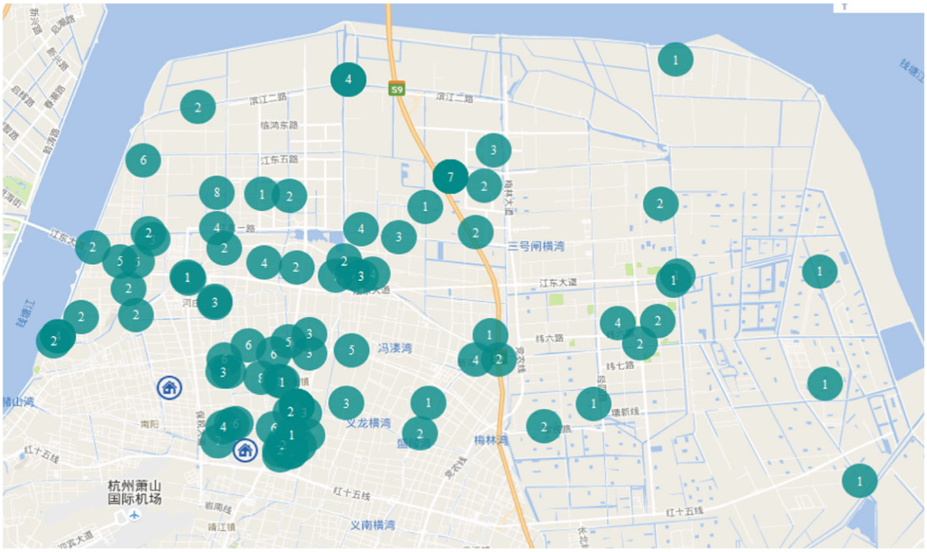


Figure 4 One working area of SGZEP, in Hangzhou city, Zhejiang Province of China. The faulty meters are marked by green color with severity level indicated. The blue houses are working stations sending out working crews to fix those faulty meters

for the working crews to consider next reparation target other than distance and locations. Combining (1) – (4), the discretized severity level formula can be written as:

$$f(x_i) = j(x_i) + s(x_i), \tag{5}$$

It can be easily seen that highest discretized severity level is 9.

4 Methodology

Considering a smart grid consisting of over 25 million smart meters with over 100,000 faulty devices existence at different locations, with limited number of working crews, an intelligent maintenance framework is always demanded taking 1) the travel distance for each working crew; 2) the maximal faulty meters that can be fixed each day; and 3) the overall aggregate severity levels that can be eliminated each day.

Figure 4 shows one working area of SGZEP, in Hangzhou city, Zhejiang Province of China. The faulty meters are marked by green color with severity level indicated. The blue houses are working stations sending out working crews to fix those faulty meters.

With the assistance of the digital map, e.g., the Google traffic map [20, 43], we connect all faulty meter locations and work stations with edges. For intersecting edges, we remove the longer ones and form a planar graph [2] (Figure 5). Two intelligent maintenance framework are proposed, which are based on adaptive genetic algorithm and customized K-medoids clustering method.

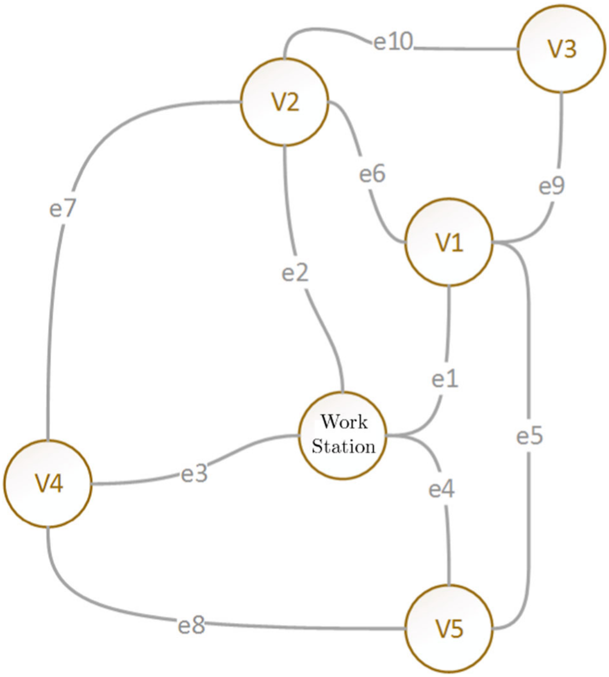


Figure 5 Planar graph connecting faulty meters and work stations

4.1 An intelligent maintenance framework based on adaptive genetic algorithm

Genetic algorithm (GA) is proposed mimicking the genetic evolution in biology, which is commonly used searching for globally optimal solution [40]. It constructs all candidate solutions as chromosomes that are formed by small pieces of genes. Starting from any candidate chromosome, GA exchange genes from that particular chromosome with crossover operations and mutation operations to search for better solutions [34]. The most fitted solution after a number of iteration of evaluations will be outputted as the optimal solution.

$$fitness(\text{path}) = \frac{\text{total traveling time} + \text{total fixing hours}}{\text{aggregate severity rates}},$$

The first intelligent maintenance framework is proposed based on an adaptive GA that allows crossover probability and mutation probability to be adjusted during the global optimal search process [25]. Starting from a particular work station, there exist multiple paths for the working crew to traverse available faulty device locations. Each of these faulty meter locations is considered as a gene; and a working path connecting the faulty meter locations is considered as a chromosome [46]. The adaptive genetic algorithm (AGA) generates new series of chromosome using crossover and mutation operations, where crossover operation searches the optimal solution globally; and mutation operation searches the optimal solution locally [35]. The fitness function for this intelligent maintenance framework is.

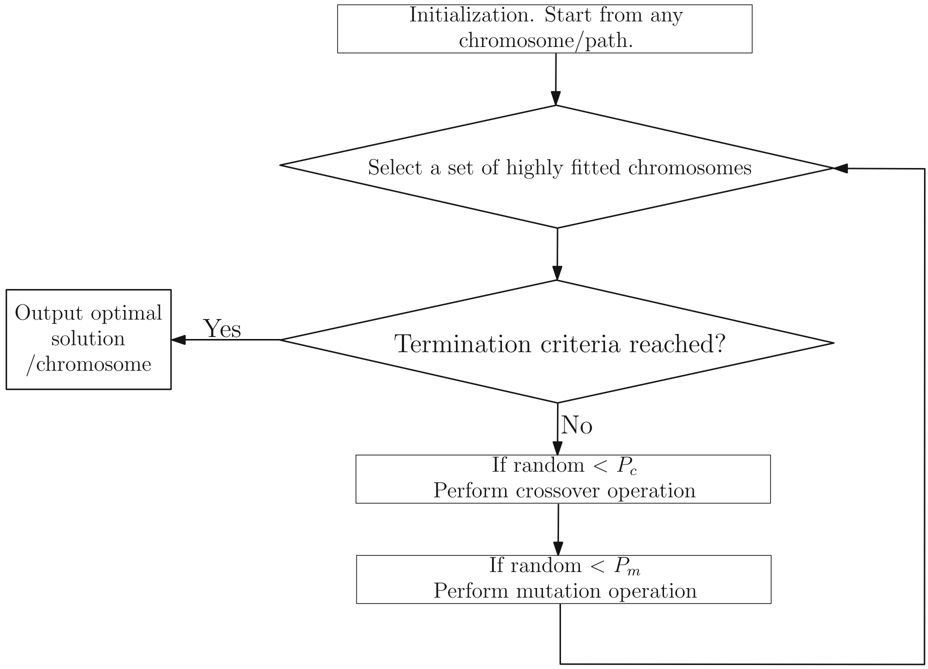


Figure 6 The adaptive genetic algorithm searching for the optimal path covering the most faulty meters with high total severity measurement rates and short travel distance

The overall AGA algorithm flowchart is depicted in Figure 6, where P_c stands for the crossover probability and P_m stands for the mutation probability. It is noted that P_c and P_m are self-adjustable according to Formulas (6) and (7) in the search process of AGA algorithm, where f, f_{max}, f_{avg} and f' represent the current fitness, maximum fitness, average fitness, and parent fitness, respectively, in the searching process. And k_1, k_2, k_3, k_4 are control variables ranged from (0,1). In this study, we set k_1, k_2, k_3, k_4 to be 0.9, 0.6, 0.1 and 0.001, respectively.

$$P_c = \begin{cases} k_1 \cdot \frac{f_{max} - f'}{f_{max} - f_{avg}} & \text{if } f' \geq f_{avg}, \\ k_2 & otherwise \end{cases} . \tag{6}$$

$$P_m = \begin{cases} k_3 \cdot \frac{f_{max} - f}{f_{max} - f_{avg}} & \text{if } f \geq f_{avg}, \\ k_4 & otherwise \end{cases} . \tag{7}$$

On the actual working map as we shown in Figure 4, there are multiple work stations that can send working crews to fix faulty meters. The first intelligent maintenance framework is proposed under this context to select the shortest working time path, eliminate the most severity levels and assign multiple workers simultaneously considering their working hours. The first proposed GA based intelligent maintenance framework (GAMF) is listed in Algorithm 1.

Algorithm 1 A genetic algorithm based intelligent maintenance framework for large-scale grid.

Input: Work stations locations and faulty meters locations.
Output: Assign working path to all available works. Or there is no more faulty meters to be assigned
Initialization: Using AGA algorithm to select one optimal working path for each work station.
While There is available working crew or faulty meter:
    Find the station with shortest working hours.
    Send out one available working crew from the found station.
    Mark all faulty meters on the working path as 'visited'.
    Update the working map.
end-While

4.2 An intelligent maintenance framework using customized k-medoids algorithm

The original K-mediods algorithm was proposed by Kaufman and Rousseeuw in 1987 [21]. The K-mediods algorithm improved the traditional K-means algorithm by placing centoids on data points instead of the actual cluster center to avoid local extreme solutions and empty clusters [32]. Moreover, the K-mediods algorithm allows the users to alter the entropy function from Euclidean distance function to others for customized design. In this study, the entropy function customized using severity levels, traveling time and fixing hours.

$$Entropy(\text{path}) = \frac{\tau \cdot \text{traveling time} + \varphi \cdot \text{fixing hours}}{\nu \cdot \text{severity rates}},$$

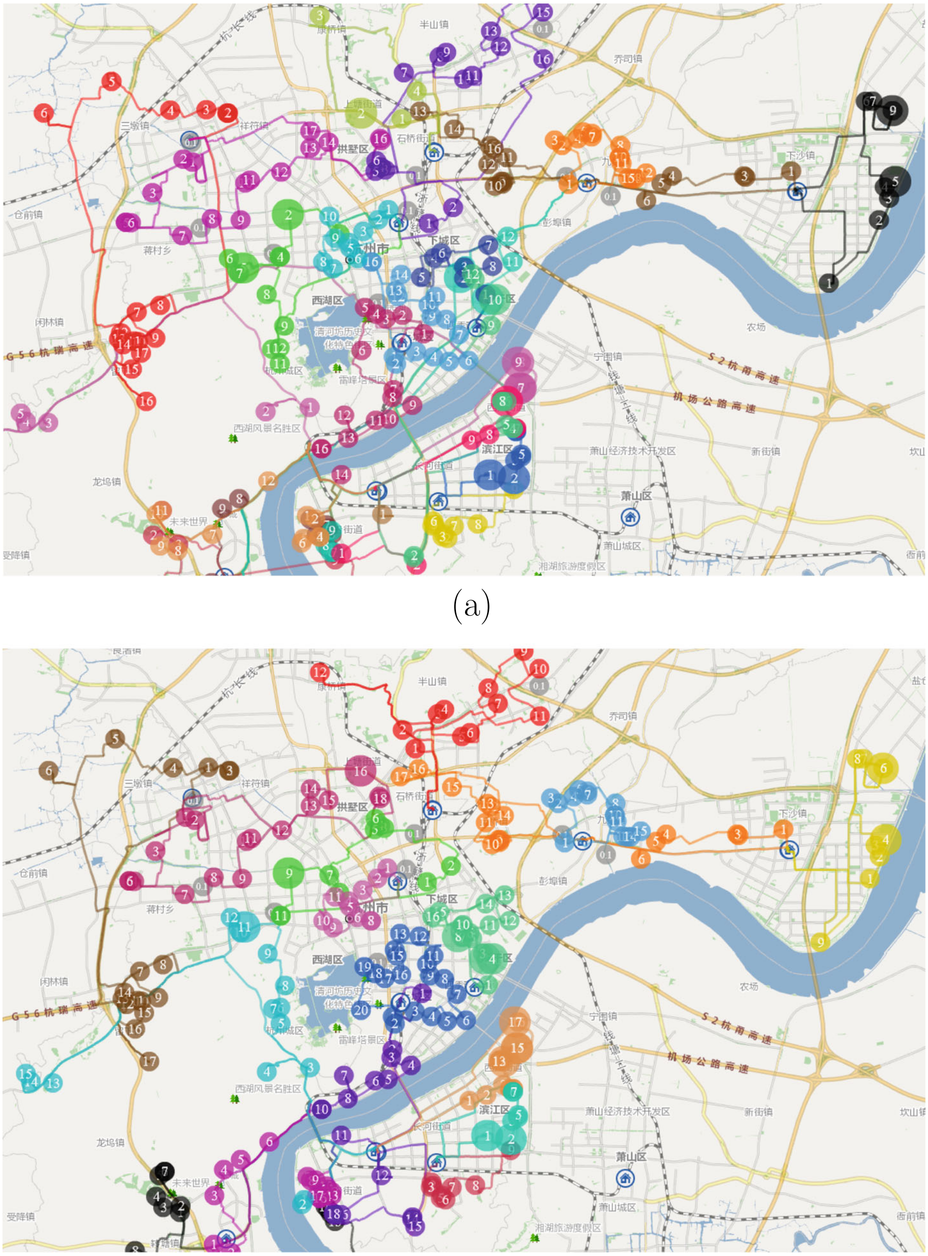
The initial values of τ, φ and ν are set as 1, 0.5 and 0.5. The optimization of the three parameters are performed in the experimental section.

The customized algorithm of K-mediods is listed in Algorithm 2, where K is the number of available working crews. The initial seeds of the K centroids are set as the locations of the highest K severity levels faulty devices' locations.

Algorithm 2 A customized K-medoids algorithm for maintaining large-scale grid.

Input: All faulty devices' locations and severity levels.
Output: K clusters, where K is the number of available working crews.
Initialization: Select the top K severity levels faulty devices as seeds.
Associate each data point to the closest centroid by customized entropy function.
While The cost of the configuration decreases:
    For each centroid c, for each non-centroid data point m:
        Swap c and m, update the graph by associating each data point to the closest centroid by the customized entropy function.
        If the total cost of the configuration increased in the previous step,
        Then undo the swapping.
        If two seeds belong to a same cluster,
        Then undo the swapping and exit the program.
    end-For
end-While

In algorithm 2, all faulty devices are sorted according to their severity level. And the top K severity faulty devices are selected as seeds. We assumed each maintenance work



(a)

(b)

Figure 7 K-medoids clustering results for faulty smart meter locations in Hangzhou city, China on Nov 12, 2018 (**a**) and Nov 13, 2018 (**b**). Fault devices marked in same color belong to the same cluster. The numbers on nodes indicate the actually reparation sequence given by the work crew starts from the initial seed position; and each working crew will be assigned to one cluster for maintenance work. The overall algorithm terminates when the K-medoids algorithm converges or any two seeds landing at the same cluster. Figure 7 depicts the clustering results of faulty smart meters located in Hangzhou city, China on Nov 12, 2018 (a) and Nov 13, 2018 (b). Fault devices marked in same color belong to the same cluster; and the numbers on nodes indicate the actually reparation sequence given by the work crew.

5 Experimental comparison and analysis

The proposed two intelligent power grid maintenance framework, i.e., genetic algorithm based maintenance framework (GAMF) and K-medoids clustering based maintenance framework (KMCMF), are compared with two trivial maintenance planning strategies named 'highest severity level first' (HSF) and 'shortest path first' (SPF) with five different datasets collected by SGZEP with five different cities/counties located in eastern China, including Hangzhou city, Cixi city, Fenghua city, Huzhou city and sanmen county. Before the experimental results are presented, we briefly describe the two trivial strategies that are only used for comparison purposes.

5.1 Highest severity level first strategy

The most intuitive way of faulty meter reparation for the maintenance working crew is to select the faulty meter with the highest severity level. The general approach of the highest severity level first strategy is listed in Algorithm 3.

Algorithm 3 Highest severity level first strategy.

1: Sort all faulty meter locations from highest severity level to lowest severity level. For same sever level locations, sort them from nearest to farthest.
2: Select the faulty meter location with highest sever level; and find the nearest working station.
3: According to the location of the selected working station and the highest sever level faulty meter, find the shortest path following Dijkstra algorithm [33].
4: Mark all faulty meters on the shortest path as 'visited'.

5.2 Shortest path first strategy

An alternative choice other than the HSF is always following the shortest path to all available faulty meters for each working crew, regardless of the severity level. Apparently, this strategy is more undesired compared with the HSF strategy. We put this strategy into comparison simply because currently, most grid companies in China are still using this simple strategy for power grid maintenance.

5.3 Results

Simulations are performed based on the data collected in five different cities located in eastern China. The dates of collection, available numbers of working crews, the numbers of faulty smart meters, averaged elapsed time for each faulty device of all five datasets are listed in Table 1.

Table 1 The dates of collection, available numbers of working crews, the numbers of faulty smart meters, averaged elapsed reparation time for each faulty device of all five datasets corresponding to the simulations on five different cities located in eastern China, including Hangzhou city, Cixi city, Fenghua city, Huzhou city and sanmen county

Dataset	Collection date	Available # of working crews	# of faulty smart meters	Averaged reparation time (min)
Hangzhou	Feb. 20, 2018	10	1062	64.2
Cixi	Jan. 8, 2019	10	628	38.3
Fenghua	Jan. 8, 2019	10	484	66.9
Huzhou	Jan. 8, 2019	10	523	49.2
Sanmen	Jan. 8, 2019	10	236	93.4

For each dataset, the simulations are performed assuming the averaged traveling speed of each working crew is either 20 km/h or 30 km/h. The maximum time of working period is set to 8 hours. The aggregate severity levels, actual working time, the aggregate travel distance of all crews and the total number of faulty devices that have been fixed using the four mentioned strategies, i.e., GAMF, KMCMF, HSF and SPF, on the five different datasets collected from five cities/counties in eastern China are shown in Tables 2, 3, 4, 5 and 6.

In summary, from Tables 2–6, all four compared methods are capable to eliminate faulty devices effectively. KMCMF eliminates the most severity levels among all datasets, followed by GAMF, HSF and SPF. HSF has relatively good performance on eliminating the overall severity levels, but has poor performance on eliminating the number of faulty devices. SPF is good at eliminating the total number of faulty devices but not taking care of the total severity levels of the fixed devices. KMCMF and GAMF take care of the both, i.e., eliminating as many severity levels as possible and concurrently traveling less distance and fixing adequate amount of faulty devices.

Table 2 The aggregate severity levels, actual working time, the aggregate travel distance of all crews and the total number of faulty devices that have been fixed using the four strategies (GAMF, KMCMF, HSF and SPF) on the dataset collected from Hangzhou city, China

Mov. speed	Strategy	Aggr. sever. rates	Total work. time (min)	Aggr. trav. dist. (km)	Total faul. dev. fixed
	KMCMF	**210**	426.94	32.63	86
20	GAMF	204	442.82	34.36	86
	HSF	199	436.92	36.31	31
	SPF	150	443.42	14.66	82
	KMCMF	**216**	442.08	29.64	87
30	GAMF	209	463.58	34.19	89
	HSF	192	473.36	47.03	25
	SPF	155	456.82	15.55	88

The proposed KMCMF eliminates the most severity levels for different moving speeds (shown in bold)

Table 3 The aggregate severity levels, actual working time, the aggregate travel distance of all crews and the total number of faulty devices that have been fixed using the four strategies (GAMF, KMCMF, HSF and SPF) on the dataset collected from Cixi city, China

Mov. speed	Strategy	Aggr. sever. rates	Total work. time (min)	Aggr. trav. dist. (km)	Total faul. dev. fixed
	KMCMF	**213**	463.00	34.90	153
20	GAMF	208	452.52	38.77	151
	HSF	200	474.16	50.89	50
	SPF	177	463.28	31.92	154
	KMCMF	**223**	463.58	40.65	165
30	GAMF	218	451.70	29.40	157
	HSF	216	438.76	67.56	56
	SPF	194	470.17	32.78	161

The proposed KMCMF eliminates the most severity levels for different moving speeds (shown in bold)

Table 4 The aggregate severity levels, actual working time, the aggregate travel distance of all crews and the total number of faulty devices that have been fixed using the four strategies (GAMF, KMCMF, HSF and SPF) on the dataset collected from Fenghua city, China

Mov. speed	Strategy	Aggr. sever. rates	Total work. time (min)	Aggr. trav. dist. (km)	Total faul. dev. fixed
	KMCMF	**444**	429.18	52.49	72
20	GAMF	431	448.36	56.75	68
	HSF	409	472.54	72.18	35
	SPF	363	463.28	31.92	72
	KMCMF	**477**	417.54	53.47	79
30	GAMF	455	464.07	59.53	75
	HSF	439	474.19	73.14	45
	SPF	373	452.29	39.49	80

The proposed KMCMF eliminates the most severity levels for different moving speeds (shown in bold)

Table 5 The aggregate severity levels, actual working time, the aggregate travel distance of all crews and the total number of faulty devices that have been fixed using the four strategies (GAMF, KMCMF, HSF and SPF) on the dataset collected from Huzhou city, China

Mov. speed	Strategy	Aggr. sever. rates	Total work. time (min)	Aggr. trav. dist. (km)	Total faul. dev. fixed
	KMCMF	**327**	381.26	77.86	99
20	GAMF	311	363.83	79.33	96
	HSF	289	386.15	87.28	38
	SPF	186	388.45	70.12	103
	KMCMF	**349**	414.47	89.98	103
30	GAMF	330	429.20	94.35	99
	HSF	318	472.37	117.63	44
	SPF	225	456.65	81.77	109

The proposed KMCMF eliminates the most severity levels for different moving speeds (shown in bold)

Table 6 The aggregate severity levels, actual working time, the aggregate travel distance of all crews and the total number of faulty devices that have been fixed using the four strategies (GAMF, KMCMF, HSF and SPF) on the dataset collected from Sanmen county, China

Mov. speed	Strategy	Aggr. sever. rates	Total work. time (min)	Aggr. trav. dist. (km)	Total faul. dev. fixed
	KMCMF	**117**	390.02	30.21	54
20	GAMF	116	414.51	28.85	51
	HSF	114	424.52	45.36	32
	SPF	112	416.25	26.77	56
	KMCMF	**121**	407.15	32.97	58
30	GAMF	118	438.30	31.10	53
	HSF	116	464.20	70.00	36
	SPF	115	454.06	34.23	59

The proposed KMCMF eliminates the most severity levels for different moving speeds (shown in bold)

The improvements on aggregate severity level of faulty devices that has been fixed can be viewed clearly from Figure 8. In overall, the proposed two intelligent maintenance framework have obvious advantage in eliminating the aggregated severity level over all faulty devices. Among the two proposed algorithms, KMCMF has slightly better performance than GAMF. GAMF is useful when a sub-optimal solution is acceptable, since it is a more efficient algorithm compared to KMCMF.

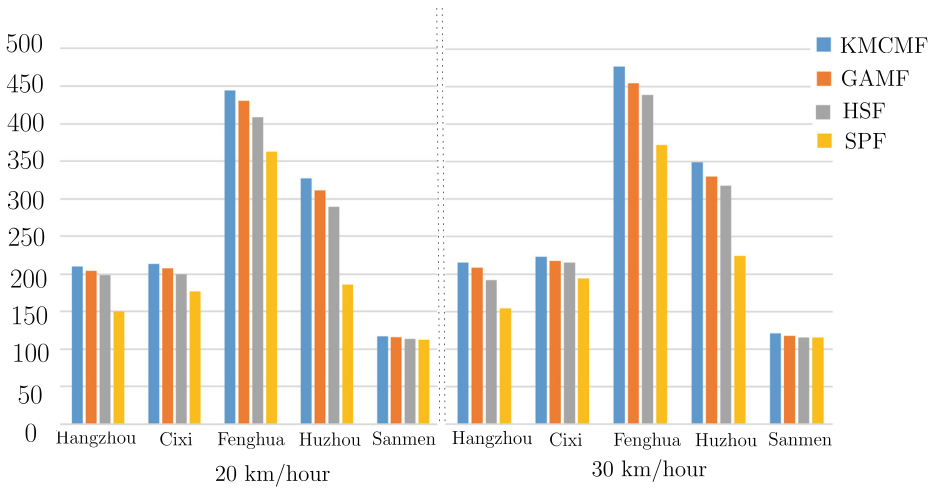


Figure 8 The improvement on eliminating aggregate severity levels with different datasets on five cities/counties in China. The results are consistent with different working crews' moving speed at 20/30 km/hour. The proposed KMCMF and GAMF has obvious advantage in eliminating the most severity levels compared with the two traditional strategies (HSF and SPF)

6 Conclusion

In this study, two intelligent maintenance frameworks are proposed for large-scale power grids. The two intelligent maintenance frameworks are built based on state-of-art AI algorithms, namely, the genetic algorithm and the K-medoids clustering algorithm. Both algorithms were extended and customized to ensure the novelty of this work. In the experiment phase, five real-world datasets collected by State Grid Zhejiang Electric Power LTD (SGZEP) from five different local cities/counties are adopted. A comparative study is performed where the two proposed intelligent maintenance frameworks are compared with the traditional maintenance planning strategies. The simulation results show the obvious advantage of the proposed algorithms over traditional approaches. The K-medoids clustering algorithm based maintenance framework (KMCMF) shows slightly better performance than the GA based maintenance framework (GAMF). The GAMF still outperforms the traditional maintenance strategies and can be useful when efficiency is the top-priority.

We summarize the main contributions of the current work to the literature:

- **More sophisticated severity level formula.** A more elegant formula calculating the severity levels for faulty devices is established. The new formula considers the three most important factors: the fault lasting hours, averaged power consumption each day for the faulty meter and the number of days towards the next billing day.
- **AI-enhanced smart maintenance methods.** We propose two smart maintenance methods using two state-of-arts AI methods. The proposed AI-enhanced methods represent the next-generation automated maintenance scheduling for large-scale smart grids and are verified to be useful based on our simulation results.
- **More comprehensive datasets for verification. Five different datasets that are collected from five different locations in China are employed. The comprehensiveness of the dataset is important to verify the effectiveness of our proposed methods.**

There are several future working directions. First of all, current maintenance strategies are proposed based on the fact that the number of faulty devices is fixed. However, in real-world scenarios, additional faulty devices may appear in the process of maintenance. A dynamic maintenance scheduling scheme is required in those cases. Second, the maintenance work may not follow exactly the assigned path in real-world scenarios. The actual working path important to improve the current maintenance scheduling system. Last, more sophisticated AI algorithms, such as reinforcement learning [36], can be adopted to further optimize the maintenance assignments.

Acknowledgements This study is supported by National Natural Science Foundation of China (No. 61850410531). It is also partially supported by Zhejiang Provincial Natural Science Foundation of China under Grant No. LY19F020016 and National Natural Science Foundation of China (No. 61602431). A previous version of this paper titled 'Data-Driven Intelligent Maintenance Planning of Smart Meter Reparations for Large-Scale Smart Electric Power Grid' was published with The 4th IEEE International Conference on Internet of People (IoP 2018) on October 8-12, 2018, at Guangzhou in China.

Compliance with Ethical Standards

Conflict of interests All authors declare that there is no conflict of interest regarding the publication of this manuscript.

References

1. Al-Jarrah, O.Y., Yoo, P.D., Muhaidat, S., Karagiannidis, G.K., Taha, K.: Efficient machine learning for big data: a review. Big Data Research **2**(3), 87–93 (2015)
2. Berge, C., Minieka, E.: Graphs and hypergraphs (1973)
3. Bhanu, B., Lee, S., Ming, J.: Adaptive image segmentation using a genetic algorithm. IEEE Trans. Syst. Man Cybern. **25**(12), 1543–1567 (1995)
4. Bishnu, P.S., Bhattacherjee, V.: Application of k-medoids with kd-tree for software fault prediction. ACM SIGSOFT Software Engineering Notes **36**(2), 1–6 (2011)
5. Bui, N., Castellani, A.P., Casari, P., Zorzi, M.: The internet of energy: a Web-enabled smart grid system. IEEE Netw. **26**(4), 39–45 (2012)
6. Cai, B., Huang, L., Xie, M.: Bayesian networks in fault diagnosis. IEEE Trans. Ind. Inf. **13**(5), 2227–2240 (2017)
7. Chen, X.W., Lin, X.: Big data deep learning: challenges and perspectives. IEEE access **2**, 514–525 (2014)
8. Colak, I.: Introduction to smart grid. In: Smart Grid Workshop and Certificate Program (ISGWCP). International, pp. 1–5, IEEE (2016)
9. Desai, B., Lebow, M.: Needed: asap approach, intelligent fleet management for transformers: a risk-based approach. IEEE Power Energ. Mag. **8**(6), 53–60 (2010)
10. Domingos, P.: A few useful things to know about machine learning. Commun. ACM **55**(10), 78–87 (2012)
11. Fan, W., Bifet, A.: Mining big data: current status, and forecast to the future. ACM sIGKDD Explorations Newsletter **14**(2), 1–5 (2013)
12. Fei, S.w., Zhang, X.b.: Fault diagnosis of power transformer based on support vector machine with genetic algorithm. Expert Systems with Applications **36**(8), 11:352–11:357 (2009)
13. Feng, J.X., Kusiak, A.: Data mining applications in engineering design, manufacturing and logistics. Int. J. Prod. Res. **44**(14), 2689–2694 (2006)
14. Germán, M.O., Molina, J.D., Romero, A.A., Gómez, H.D., García, E.: Power asset management: methods and experiences in colombian power system. In: Transmission & Distribution Conference and Exposition-Latin America (PES T&D-LA), 2014 IEEE PES, pp. 1–6, IEEE (2014)
15. Ijjina, E.P., Chalavadi, K.M.: Human action recognition using genetic algorithms and convolutional neural networks. Pattern Recognit. **59**, 199–212 (2016)
16. Isermann, R.: Fault-diagnosis applications: model-based condition monitoring: actuators, drives, machinery, plants, sensors, and fault-tolerant systems. Springer Science & Business Media (2011)
17. Jardine, A.K., Lin, D., Banjevic, D.: A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech. Syst. Signal Process. **20**(7), 1483–1510 (2006)
18. Jiang, H., Zhang, J.J., Gao, W., Wu, Z.: Fault detection, identification, and location in smart grid based on data-driven computational methods. IEEE Trans. Smart Grid **5**(6), 2947–2956 (2014)
19. Jordan, M.I., Mitchell, T.M.: Machine learning: Trends, perspectives, and prospects. Science **349**(6245), 255–260 (2015)
20. Juntunen, T., Kostakos, V., Perttunen, M., Ferreira, D.: Web tool for traffic engineers: direct manipulation and visualization of vehicular traffic using google maps. In: Proceeding of the International Academic Mindtrek Conference, pp 209–210 (2012)
21. Kaufman, L., Rousseeuw, P.: Clustering by means of medoids. North-Holland (1987)
22. Lee, J., Ni, J., Djurdjanovic, D., Qiu, H., Liao, H.: Intelligent prognostics tools and e-maintenance. Comput. Ind. **57**(6), 476–489 (2006)
23. Lei, R., Xue, L., Le, X., Liu, W.: Coordinated energy cost management of distributed internet data centers in smart grid. IEEE Trans. Smart Grid **3**(1), 50–58 (2012)
24. Lei, Y., He, Z., Zi, Y., Chen, X.: New clustering algorithm-based fault diagnosis using compensation distance evaluation technique. Mech. Syst. Signal Process. **22**(2), 419–435 (2008)
25. Lu, H., Chen, J., Yan, K., Jin, Q., Xue, Y., Gao, Z.: A hybrid feature selection algorithm for gene expression data classification. Neurocomputing **256**, 56–62 (2017)
26. Mühlenbein, H., Schomisch, M., Born, J.: The parallel genetic algorithm as function optimizer. Parallel Comput. **17**(6-7), 619–632 (1991)
27. Najafabadi, M.M., Villanustre, F., Khoshgoftaar, T.M., Seliya, N., Wald, R., Muharemagic, E.: Deep learning applications and challenges in big data analytics. Journal of Big Data **2**(1), 1 (2015)
28. Nelson, V.P.: Fault-tolerant computing: Fundamental concepts. Computer **23**(7), 19–25 (1990)
29. Qin, S.J.: Survey on data-driven industrial process monitoring and diagnosis. Annu. Rev. Control. **36**(2), 220–234 (2012)
30. Rai, A., Upadhyay, S.: Bearing performance degradation assessment based on a combination of empirical mode decomposition and k-medoids clustering. Mech. Syst. Signal Process. **93**, 16–29 (2017)

31. Siewiorek, D.P., Swarz, R.S.: The theory and practice of reliable system design. Digital press (1982)
32. Singh, S.S., Chauhan, N.: K-means v/s k-medoids: A comparative study. In: National Conference on Recent Trends in Engineering & Technology, vol. 13 (2011)
33. Skiena, S.: Dijkstra algorithm. Implementing Discrete mathematics: Combinatorics and Graph Theory with Mathematica, pp. 225–227. Addison-Wesley, Reading (1990)
34. Spears, W.M.: Crossover or mutation? In: Foundations of Genetic Algorithms, vol 2. pp 221–237, Elsevier (1993)
35. Srinivas, M., Patnaik, L.M.: Adaptive probabilities of crossover and mutation in genetic algorithms. IEEE Trans. Syst. Man Cybern. **24**(4), 656–667 (1994)
36. Sutton, R.S., Barto, A.G.: Reinforcement learning: An introduction. MIT press (2018)
37. Venkatasubramanian, V., Rengaswamy, R., Kavuri, S.N., Yin, K.: A review of process fault detection and diagnosis: Part iii: Process history based methods. Comput. Chem. Eng. **27**(3), 327–346 (2003)
38. Viswanadham, N., Johnson, T.: Fault detection and diagnosis of automated manufacturing systems. In: Proceedings of the 27th IEEE Conference on Decision and Control, 1988, pp. 2301–2306. IEEE (1988)
39. Wang, W., Chen, H., Lou, B., Jin, N., Lou, X., Yan, K.: Data-driven intelligent maintenance planning of smart meter reparations for large-scale smart electric power grid. In: 2018 IEEE Smartworld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), pp. 1929–1935. IEEE (2018)
40. Whitley, D.: A genetic algorithm tutorial. Stat. Comput. **4**(2), 65–85 (1994)
41. Witten, I.H., Frank, E., Hall, M.A., Pal, C.J.: Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann (2016)
42. Wu, X., Zhu, X., Wu, G.Q., Ding, W.: Data mining with big data. IEEE Trans. Knowl. Data Eng. **26**(1), 97–107 (2014)
43. Wu, Y.J., Wang, Y., Qian, D.: A google-map-based arterial traffic information system. In: Intelligent Transportation Systems Conference, 2007. ITSC 2007, pp. 968?-973. IEEE (2007)
44. Yan, K., Ji, Z., Lu, H., Huang, J., Shen, W., Xue, Y.: Fast and accurate classification of time series data using extended elm: Application in fault diagnosis of air handling units. IEEE Transactions on Systems, Man, and Cybernetics: Systems (2017)
45. Yan, K., Zhong, C., Ji, Z., Huang, J.: Semi-supervised learning for early detection and diagnosis of various air handling unit faults. Energ. Buildings **181**, 75–83 (2018)
46. Yang, J., Wu, C., Lee, H.P., Liang, Y.: Solving traveling salesman problems using generalized chromosome genetic algorithm. Prog. Nat. Sci. **18**(7), 887–892 (2008)
47. Zhong, C., Yan, K., Dai, Y., Jin, N., Lou, B.: Energy efficiency solutions for buildings: Automated fault diagnosis of air handling units using generative adversarial networks. Energies **12**(3), 527 (2019)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.