energies

MDPI

Article

Defect Data Association Analysis of the Secondary System Based on AFWA-H-Mine

Yan Xu, Mingyu Wang * and Wen Fan

State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, North China Electric Power University (Baoding), Baoding 071003, China; 51351343@ncepu.edu.cn (Y.X.); 2192213143@ncepu.edu.cn (W.F.)

* Correspondence: 2192213236@ncepu.edu.cn; Tel.: +86-177-2172-9598 (ext. 071003)

Abstract: The fault data of the secondary system of smart substations hide some information that the association analysis algorithm can mine. The convergence speed of the Apriori algorithm and FP-growth algorithm is slow, and there is a lack of indicators to evaluate the correlation of association rules and the method to determine the parameter threshold. In this paper, the H-mine algorithm is used to realize the fast mining of fault data. The algorithm can traverse data faster by using the data structure of the H-struct. This paper also sets the lift and CF value to screen the association rules with good correlation. When setting the three key parameters of association analysis, namely, support threshold, confidence threshold, and lift threshold, an objective function composed of weighted average lift, CF value, and data coverage rate was selected, and the adaptive fireworks algorithm was used to optimize the parameters in the association analysis. In particular, the rule screening strategy is introduced in fault cause analysis in this paper. By eliminating rules with high similarity, derived signals in association rules are eliminated to the greatest extent to improve the readability of rules and ensure easy understanding of results.

Keywords: fault analysis of secondary system of smart substation; AFWA; H-mine; parameter optimization of association rules

Citation: Xu, Y.; Wang, M.; Fan, W. Defect Data Association Analysis of the Secondary System Based on AFWA-H-Mine. *Energies* **2021**, *14*, 4228. https://doi.org/10.3390/en14144228

Received: 7 June 2021
Accepted: 9 July 2021
Published: 13 July 2021

1. Introduction

In a substation, the secondary system plays a pivotal role in the control, protection, and regulation of the primary equipment, and the reliable operation of the secondary system is related to the safety and stability of the power system [1,2].

Smart substations adopt advanced intelligent equipment, realize the digitization and information sharing of the whole station, and can automatically collect operating data [3,4]. Therefore, in a smart substation, a large amount of data related to the running state of secondary equipment can be obtained, including online operation information, historical defect information, fault record, protection action information, etc. [5]. The machine learning model can be used to mine the information behind these data [6], which is helpful for maintenance personnel to check the working state of the secondary equipment and repair the weak part in the secondary equipment before any abnormal situation occurs in the secondary equipment.

This hidden information usually cannot use simple mathematical statistics for analysis. At present, a lot of machine learning methods are applied to the data analysis of the secondary system, one paper [7] used deep learning implementation of fault prediction, while another paper [8] used the binary chart correlation model combined with the Bayesian suspected degree for the calculation of the fault probability. However this latter method can only be used for predictions; it cannot realize the mining of association information.

At present, the association analysis algorithm is also applied to the analysis of defect data of the secondary system. The association analysis algorithm can reflect the hidden information between data by using association rules. One paper [9,10] used the Apriori

algorithm, and another [11] used the FP-Growth algorithm to analyze the hidden information inside the defect data of the secondary system of the smart substation. However, when these algorithms generate frequent itemsets, they all need to scan the dataset several times, which will increase the time required for association analysis.

To improve the efficiency of association analysis, this paper uses the H-mine [12] algorithm to mine frequent item sets. This algorithm uses the H-struct structure to process data and only mines one partition at a time. Compared with the traditional association analysis algorithm, it does not need to generate many candidate itemsets like the Apriori algorithm. Additionally, it does not need to generate FP-tree and the iterative database like the FP-Growth algorithm. In the case of sparse data, this algorithm can shorten the time of association analysis.

The traditional association analysis algorithm requires humans to determine parameters such as confidence and support; this process is time-consuming [13]. To solve this problem, optimization algorithms can calculate these parameters; one paper [14] used the particle swarm optimization (PSO) algorithm, while another paper [15] used the ant lion algorithm. However, these algorithms are greatly influenced by the initial value and are easy to converge prematurely. Because, in the traditional group algorithm, the individual behavior of the group is simple, while the difference between the individuals is poor, and there is no central control of the individual.

To solve this problem, one paper [16] proposed a fireworks algorithm, which sets different resources for individuals with different fitness to improve the overall searchability. However, this algorithm only depends on the fitness difference between individuals to determine the number and explosion radius of explosive sparks, which will make the explosion radius of fireworks with good fitness too small, leading to the poor local convergence ability of the algorithm [17]. To improve the local convergence ability of the algorithm, the AFWA [18] algorithm is used in this paper to improve the local convergence ability of the algorithm. This algorithm can effectively adjust the step size according to the search situation, so the local convergence ability is better than the fireworks algorithm.

In summary, the H-Mine algorithm is used in this paper to mine frequent item sets. Furthermore, the adaptive fireworks algorithm is combined to optimize the selection of key parameters and mines hidden association rules in defect records of the secondary system. Thus, the method proposed in this paper adjusts threshold parameters of the association analysis, is less dependent on humans, and simplifies the analysis process. However, the method proposed in this paper only involves the process of data analysis, without the process of data processing, so it is only applicable when high-quality defect data are present.

2. Secondary Device Defect Database Model

Smart substation operation and maintenance personnel usually record the name of the substation, the name of the equipment, the alarm signal, the date of the equipment fault, the cause of the fault, and other information when the smart substation shows abnormal functioning. This information can be obtained from the production management system or the defect log at the site.

Compared with the data in the production management system, the data at the substation site are more complete, so the data in this paper mainly comes from the defect records at the station substation site.

Generally, recorded data can be divided into the categories described in Table 1, and different categories of data can be distinguished by setting different codes.

Table 1. Data sheet.

Category	Information
Smart station information	Name of substation
Equipment information	Type of equipment and manufacturer
Fault information	Alarm signal, fault cause, and position
Event processing information	Processing conditions

The names of the substations in the database include the names of six smart substations in a certain area. The manufacturer is the actual manufacturer of the secondary equipment in these substations.

The equipment includes more than 10 kinds of equipment, including communication links, merging units, switches, protection devices, fault records, etc.

The fault causes include "program error," "configuration error," "power fault," "optical fiber fault," etc., and the fault location refers to the specific fault location of the device.

The treatment situation is the actual operation of maintenance personnel in the process of fault repair.

3. Fireworks Algorithm

3.1. Swarm Intelligence Algorithm

In recent years, the swarm intelligence algorithm has become a research hotspot, in which scholars have invested a lot of energy; moreover, they have put forward many related algorithms. For example, one paper [19] proposed the particle swarm optimization algorithm, which attracted the attention of many scholars as soon as it was proposed. Subsequently, some scholars proposed some relevant improvement strategies [20]. Still, in the traditional swarm intelligence algorithm, there is no difference between the group's individuals, and their behaviors are completely the same. Hence, the search performance is not good in many cases.

One paper [16] proposed the fireworks algorithm, which sets different resources for individuals with different fitness to improve the overall searchability. However, the following problems still exist:

1 In general, the explosion radius of fireworks with small fitness is close to 0, which leads to the lack of searchability for the optimal fireworks.
2 The explosion bias of fireworks is the same in any dimension, which will reduce the diversity of sparks.
3 When the spark goes beyond the boundary, it will be mapped to a point very close to the origin, making it difficult for the spark to find the optimal value.

To solve the above problems, an enhanced fireworks algorithm was proposed in [17], and the explosion radius, spark generation mode, mapping rules, and selection strategy of the algorithm were improved. However, there is still a problem of poor local convergence. This is because the calculation of explosion radius completely relies on the difference in fitness between fireworks without considering other information in the solution process.

To improve the local convergence ability of the algorithm, the AFWA [18] algorithm is used in this paper to improve the local convergence ability of the algorithm. This algorithm can effectively adjust the step size according to the search situation, and its local searchability is better than that of the fireworks algorithm.

3.2. Traditional Fireworks Algorithm

The firework algorithm has better global searchability. Its fitness value can balance the searchability and consider both the local and the global scale when allocating resources and information exchange, making it suitable for optimizing multi-objective problems. Moreover, the explosion mechanism can improve the accuracy and speed of the algorithm.

The calculation formula for the number of sparks and the explosion radius of a firework explosion is as follows:

$$S_i = M \times \frac{f_{\max} - f(x_i) + \varepsilon}{\sum_{i=1}^{N} (f_{\max} - f(x_i)) + \varepsilon} \tag{1}$$

$$A_i = \widehat{A} \times \frac{f(x_i) - f_{\min} + \varepsilon}{\sum_{i=1}^{N} (f(x_i) - f_{\min}) + \varepsilon} \tag{2}$$

where f_{min} and f_{max} are the minimum and maximum fitness values, respectively, A and M are constants, which are used to adjust the explosion radius and the number of fireworks, and ε is a constant that keeps the denominator from becoming 0.

In order to limit the number of sparks, it is necessary to limit the number of sparks, where the upper limit is S_{max} and the lower limit is S_{min}.

In the fireworks algorithm, several dimensions of fireworks are randomly selected for displacement, and the formula is as follows:

$$\Delta x_i^k = x_i^k + rand(0, A_i) \tag{3}$$

where $rand(0, A_i)$ is a random number in the range of $(0, A_i)$.

In addition to the explosion sparks in the firework algorithm, mutation sparks are calculated by the Gaussian mutation operator to improve the diversity of sparks:

$$x_i^k = x_i^k \times Gaussian(1, 1) \tag{4}$$

Sparks beyond the feasible range need to be remapped to a new location:

$$x_i^k = x_{\min}^k + x_i^k mod(x_{\max}^k - x_{\min}^k) \tag{5}$$

where $x_{\max}^k$ and $x_{\min}^k$ are the upper and lower limits of the solution space.

3.3. Adaptive Fireworks Algorithm

This paper uses the adaptive fireworks algorithm to realize parameter optimization to improve the search performance of the optimization algorithm used in this paper. The algorithm has better search performance than the fireworks algorithm. Compared with the fireworks algorithm, the specific improvements of the adaptive fireworks algorithm are as follows:

1. In the traditional fireworks algorithm, the explosion radius of fireworks that has small adaptability will be relatively small. In order to avoid this problem, the algorithm sets the minimum explosion radius; when $A_{ik} < A_{min,k}$, the explosion radius of firework i in dimension k is:

$$A_{ik} = A_{\min,k} \tag{6}$$

In other cases, the explosion amplitude remains unchanged.

The minimum explosion radius is selected using the following formula:

$$A_{\min,k}(t) = A_{init} - \frac{A_{init} - A_{final}}{evals_{\max}} \sqrt{(2evals_{\max} - t)t} \tag{7}$$

where $evals_{max}$ is the maximum number of evaluation times, A_{init} is the initial detection value of the explosion radius, and A_{final} is the final detection value of the explosion radius.

2. The mutation operation of the firework algorithm is enhanced to avoid the Gaussian mutation in the traditional firework algorithm that will cause too many sparks near

the origin. Moreover, the mutation between the current solution and the current optimal solution is performed:

$$x_{ik} = x_{ik} + (x_{Bk} - x_{ik}) \times e \tag{8}$$

where e is a random variable with a mean equal to 0 and a variance equal to 1, x_{ik} is the current solution, and x_{Bk} is the optimal solution in the current population.

At the same time, the mapping rules are changed to:

$$x_{ik} = x_{lb,k} + U(0,1) \cdot (x_{ub,k} - x_{lb,k}) \tag{9}$$

where $x_{ub,k}$ is the upper limit of the solution space and $x_{ib,k}$ is the lower limit of the solution space.

3. When selecting the next generation of fireworks, the traditional fireworks selection method needs to construct a Euclidean distance matrix in each generation population, which will lead to the increase of time consumption of the traditional fireworks algorithm. To avoid this problem, the adaptive fireworks algorithm first selects the individuals with the best fitness in the population as the next generation of fireworks, and then randomly selects the rest of the fireworks.
4 In the traditional fireworks algorithm, the optimal fireworks explosion radius is 0, which means that the optimal fireworks contribution to the convergence process is limited. Still, because it generates the largest number of individuals, it is of great significance for the whole convergence process, so the optimal fireworks also need to set the explosion radius.

The AFWA algorithm uses the generated sparks and the parent to find the optimal firework explosion radius in the children, respectively.

Firstly, an individual needs to be selected, and the distance between it and the optimal fireworks is used as the next generation explosion radius. The individual needs to meet the following two conditions simultaneously: the fitness is worse than that of the previous generation, and the distance between the individual and the optimal individual is the smallest. The formula is as follows:

$$\widehat{s} = \underset{s_i}{\operatorname{argmin}}(d(s_i, s^*)) \tag{10}$$

At the same time:

$$f(s_i) > f(X) \tag{11}$$

where s_i indicates the individual generated by the fireworks, s^* denotes the individual with the smallest fitness value in the current race, d refers to the distance function (in this paper, the infinite norm is used as the distance metric), and X represents the fireworks.

The explosion radius of the optimal firework is as follows:

$$A_{CF}(g+1) = \frac{\widehat{s}(g) + \widehat{s}(g+1)}{2} \tag{12}$$

where $\widehat{s}(g)$ and $\widehat{s}(g+1)$ are the shortest distances of generations g and $g + 1$, respectively.

4. Association Analysis Algorithm

4.1. Association Rule Evaluation Index

In traditional association rules, the indicators to measure the quality of rules mainly include the support threshold and confidence threshold:

$$Sup(A \to B) = \frac{Sup(BA)}{|D|} \tag{13}$$

$$Conf(A \rightarrow B) = \frac{Sup(BA)}{Sup(A)} \tag{14}$$

where D is the total amount of data, *Sup(BA)* is the amount of data containing both B and A, and *Sup(A)* is the number of occurrences of A.

Traditional association rules do not introduce other indicators to exclude association rules with poor correlation and independence. Consequently, a considerable number of meaningless association rules and even misleading association rules may be produced. To exclude useless association rules, the traditional screening indicators are:

(1) *CF:*

If *Conf(A→B) > Sup(B)*:

$$CF = \frac{Conf(A \rightarrow B) - Sup(B)}{1 - Sup(B)} \tag{15}$$

on the contrary:

$$CF = \frac{Conf(A \rightarrow B) - Sup(B)}{Sup(B)} \tag{16}$$

When *CF* is a positive number, it means that the front and back parts positively correlate. When *CF* is a negative number, it means that the two are in a negative correlation. At the same time, the closer the *CF* is to 1, the higher the confidence of the rule.

(2) *Lift:*

$$Lift(A \rightarrow B) = \frac{Sup(BA)}{Sup(A) \times Sup(B)} \tag{17}$$

The index of *lift* reflects the correlation of two variables. It is generally believed that the higher the degree of *lift*, the more obvious the positive correlation between the two variables. Taking 1 as the boundary, when the degree of *lift* is 1, there is no correlation between the two variables.

(3) The number of items in association rules N and the total number of association rules Num:

$$N(A \rightarrow B) = A + B \tag{18}$$

where A and B represent the number of prefixes and suffixes.

In general, the less the total number of prefixes and suffixes of association rules is, the more convenient it will be to understand. Meanwhile, the total number of rules can also reflect the complexity of analysis results. When the number of rules is small, it is easier for analysts to extract useful conclusions.

The above indexes are only for single association rules. The main optimization goal of this paper is to target all association rules generated by the association analysis algorithm. Therefore, when considering the above indexes, the average value of all rules should be obtained.

In addition, to prevent the mining results from not covering all the data due to too few rules, this paper also sets the index of data coverage, which refers to the proportion of the data covered by all the rules generated. The larger the index is, the more comprehensive the information mined will be. The calculation formula is as follows:

$$CR = \frac{\sum_{i=1}^{n} C_{AR_i}}{|D|} \tag{19}$$

where C_{ARi} is the transaction is covered by the i-th association rule and D is the total amount of data of a particular type in the dataset.

The parameters to be optimized in this article are: support threshold, confidence threshold, and *lift* threshold. Usually, if the threshold is set too high, the data coverage of the association rules will be reduced, and if the threshold is set too low, the results will

contain a lot of useless rules. The optimization objective is to make all association rules have a better *lift* and *CF* than possible under ensuring large data coverage. Therefore, the fitness function is as follows:

$$f(x_i) = -(W_1 * CoverRate + W_2 * (Lift_{ave} + CF_{ave})) \quad (20)$$

where W_1 and W_2 are the weights of the data coverage and the sum of *lift* and *CF*, respectively.

Among them, the possible value of *lift* is large, so it needs to be standardized:

$$Lift'_i = \frac{Lift_i - Lift_{\min}}{Lift_{\max} - Lift_{\min}} \quad (21)$$

4.2. Association Analysis Algorithm

In the era of big data, people hope to convert massive data into specific information, and data mining can uncover the hidden relationship between the data through a series of detailed analyses. This paper uses association analysis in data mining to find out the hidden relationship between the defect data of secondary equipment.

In this paper, H-mine is used as the main algorithm of association analysis. Compared with the traditional association analysis algorithm, this algorithm is faster, and its principle is as follows:

H-mine processes data by generating H-struct and only mines one partition at a time when processing data. Compared with the traditional Apriori algorithm, H-mine does not need to generate many candidate item sets, traverses data faster, and mines frequent itemsets faster. In addition, compared with FP-growth, H-mine does not generate FP-tree and the iterative database required to generate FP-tree. When the amount of data is large, the H-mine algorithm can save time compared with the FP-growth algorithm.

The following is a specific example to illustrate the specific steps of the H-Mine algorithm. The minimum support count is set to 2. The transaction set of known database TDB and the filtered frequent items are shown in Table 2.

Table 2. The transaction set.

ID	Transaction	Frequent Items
1	A, B, D, I, F	A, B, D, F
2	A, B, C, D, Z	A, B, C, D
3	B, C, E, F, K	B, C, E, F
4	B, D, E, W	B, D, E

Frequent itemsets to be mined can be divided into containing item A, containing item B, but not including item A; containing item C but not including item B and item A; containing item D but not including item A, B, C; containing item E but not including item A, B, C, and D; and only containing item F.

First, the database is scanned to find the one-itemsets required for association analysis according to the support. These itemsets are stored in the Header table H in alphabetical order. The occurrence times of them are recorded at the same time: {A: 2, B: 4, C: 2, D: 3, E: 2, F: 2}.

Then, the database is scanned again, and each item in Header Table H is used as the head pointer to connect the transactions with the same first item into a link to form an H-struct. There are three parts in Header Table H, namely the name of the transaction, the support count, and the pointer, as shown in Figure 1: after the H-struct is established, data mining is performed only on the H-struct, traverses the A-queue, finds out all the frequent items in the A-queue, establishes the Header table HA, and records the support count according to the elements in the A-queue. The output frequent 2-item set can be obtained as follows: {AB: 2, AD: 2}.

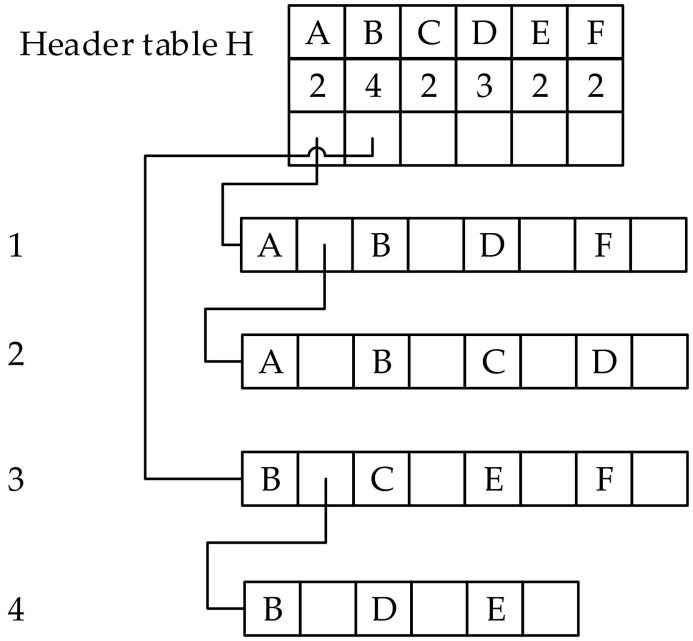


Figure 1. Header table H.

Similarly, we continue to dig out the item whose first item is AB and build Header table HA, as shown in Figure 2. It can be found that only ABD meets the requirements. Since there are no data with AD as the first item, there is no need to establish a queue about AD, so there is no need to mine AD specifically.

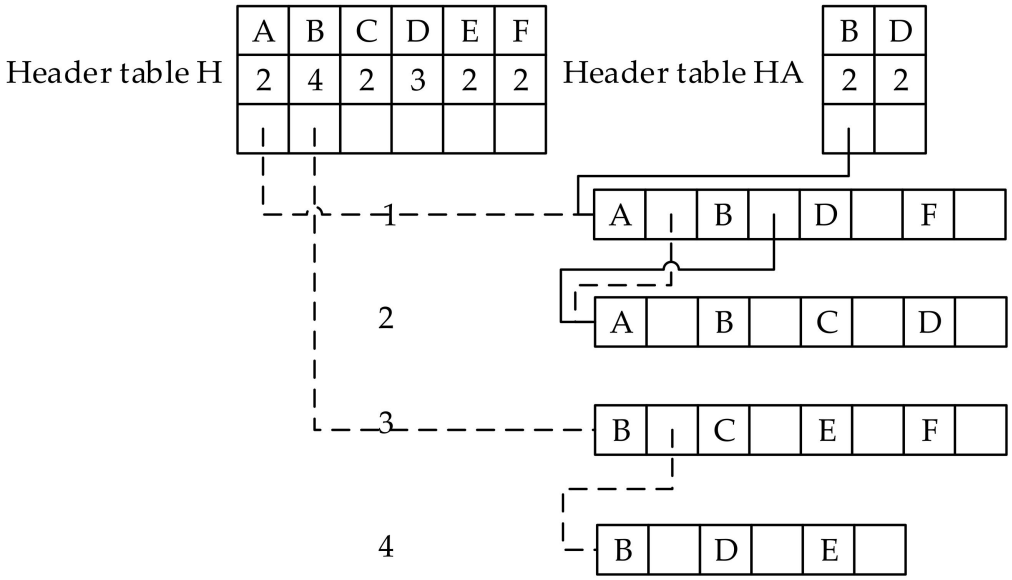


Figure 2. Header table HB.

The next step is to mine the frequent item set containing B element but not containing A element, which not only needs to mine the established queue with B element as the first item and includes the frequent items containing B element in the Header table HA in the previous step. Since the queue with AB item as the first item has been established in the Header table HA in the last step, they can be inserted into the B queue, and the result can be obtained after mining. The specific steps are the same as when mining the frequent item set containing item A, that is, {BC: 2, BD: 3, BE: 2, BF: 2}, and there is no need to process other elements in Header table H, because there is no queue that starts with them.

To sum up, the frequent item set generated by the example is: 1—item set: {A: 2, B: 4, C: 2, D: 3, E: 2, F: 2}; 2—item set: {AB: 2, AD: 2, BC: 2, BD: 3, BE: 2,BF: 2}; 3—item set: {ABD: 2}.

After all the frequent itemsets are obtained, association rules' confidence and *lift* are calculated using the frequent item sets. Then, the association rules that do not meet the confidence threshold and life threshold are deleted. The confidence and *lift* threshold need to be calculated by the adaptive fireworks algorithm.

4.3. Association Rule Screening Strategy

When analyzing the cause of a fault, alarm signals are usually accompanied by derived alarm signals, which will result in the generated association rules containing too many items, thus affecting the comprehensibility. To reduce redundant items in rules, the similarity index is set in this paper to eliminate redundant rules:

$$Simi[i,j] = \frac{S(AR_i, AR_j)}{S(AR_i)} \tag{22}$$

where $S(AR_i,AR_j)$ is the number of concurrent transactions in rules *ARi, ARj,* $S(AR_i)$ is the total number of AR_i transactions, and *Simi[i, j]* refers to the similarity of rule *i* to rule *j*.

The similarity set in this paper mainly refers to the similarity of prefixes:

$$Simi_{ant}[i,j] = \frac{S(A_i, A_j)}{S(A_i)} \tag{23}$$

In the same way, we can get the similarity of *j* to *i*. When the suffix of two rules is the same, and the prefix similarity is greater than the index, the rules with more items in the rules are filtered out.

The specific process of obtaining association rules is as follows:

1. Initialize the population.
2. Import defect data.
3. Bring the individual fireworks data into the association analysis for analysis and evaluation.
4. Determine the number of explosive sparks, core fireworks, and non-core firework explosion radius.
5. Displacement operation is carried out on the individual fireworks, and the cross-border sparks are processed.
6. Choose the next generation of fireworks.
7. Determine whether the rule as a whole satisfies the termination condition of iteration at this time. If not, return to Step 3.

In particular, when analyzing the cause of fault, an association rule screening strategy is also needed to exclude rules with a large number of items.

5. Results and Discussion

5.1. Analysis of the Frequent Item Set Mining Results

The experimental data in this paper are taken from the defect records of smart substations in a certain area in the past few years. Before mining association rules, frequent itemsets should be mined first. In this paper, frequent itemsets are mined in all three cases, which can provide the maintenance personnel with the support of defect data so that the maintenance personnel can analyze the defects with high support first. It is worth mentioning that the support level needs to be set when mining frequent item sets, so there is no need to generate association rules for frequent item set mining with the H-mine algorithm, nor to use the AFWA algorithm for optimization.

First of all, the manufacturer's data are extracted, and the equipment produced by the manufacturer is statistically prone to fault to remind the acceptance personnel to pay

attention to the equipment produced by the manufacturer when accepting the equipment. The top five manufacturers of support are shown in Figure 3.

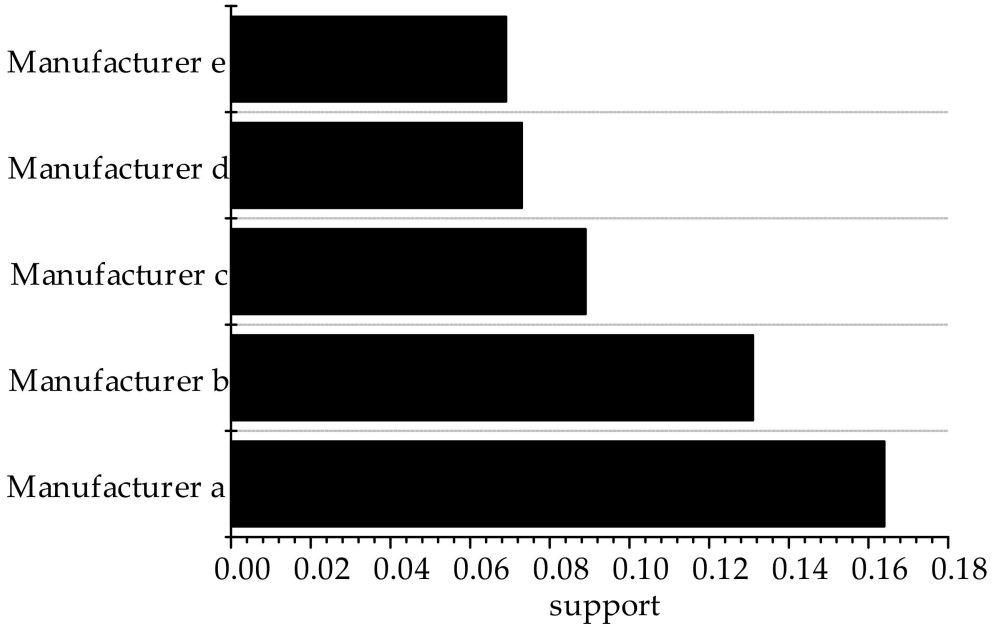


Figure 3. Manufacturer support.

As can be seen from Figure 3, manufacturers prone to equipment quality problems include A, B, C, D, and E, and their sum of support exceeds 0.5. Therefore, when substation workers check and accept the equipment produced by these manufacturers, acceptance standards should be raised.

Next, alarm signals are extracted. The frequent item set of these data can be used to count the number of alarm signals when secondary equipment faults occur in these substations to remind maintenance personnel which alarm signals should be investigated. Some of the results are shown in Figure 4.

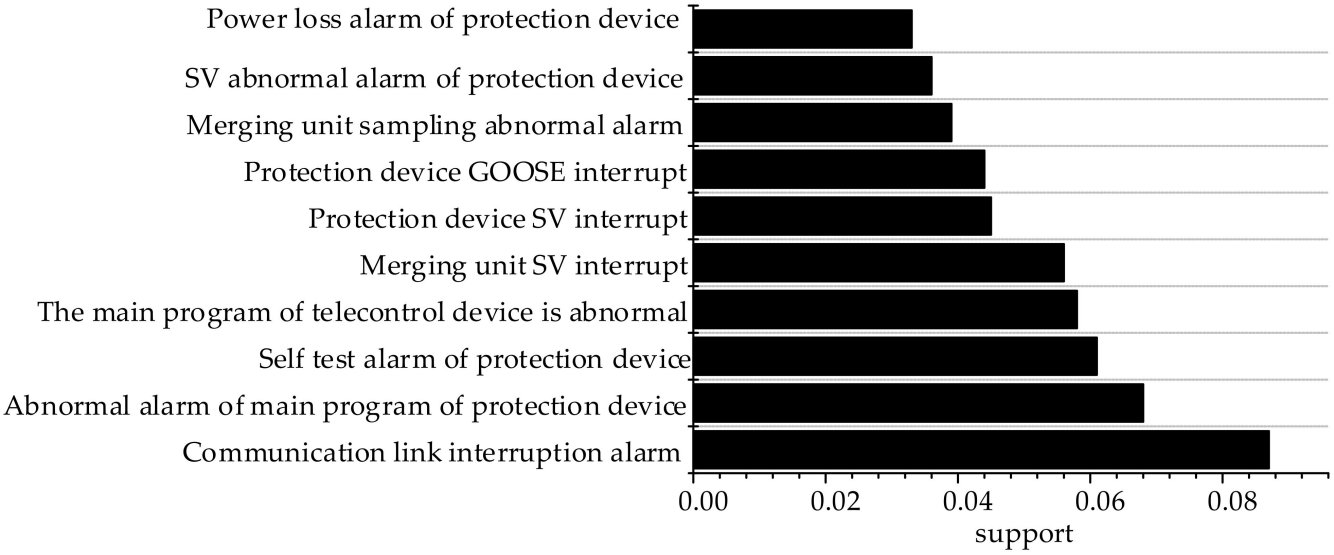


Figure 4. Alarm signal support.

As shown in Figure 4, alarm signals with a high frequency include a communication link interruption alarm, abnormal alarm of the main program of protection device, self-inspection alarm of protection device, etc. Among them, the communication link interruption alarm has the highest support. In addition, the related alarm signals of the protection device also have a high support. Therefore, maintenance personnel should focus on monitoring the reliability of the protection device and the operation of the communication link.

Finally, two types of data, the device name and fault location, are extracted and imported into the H-Mine algorithm to obtain their frequent item sets, which are used to count the defects of secondary equipment in these substations to find out which equipment is more prone to abnormal situations compared with other equipment. Some frequent itemsets with high support are selected in this paper, as shown in Figure 5.

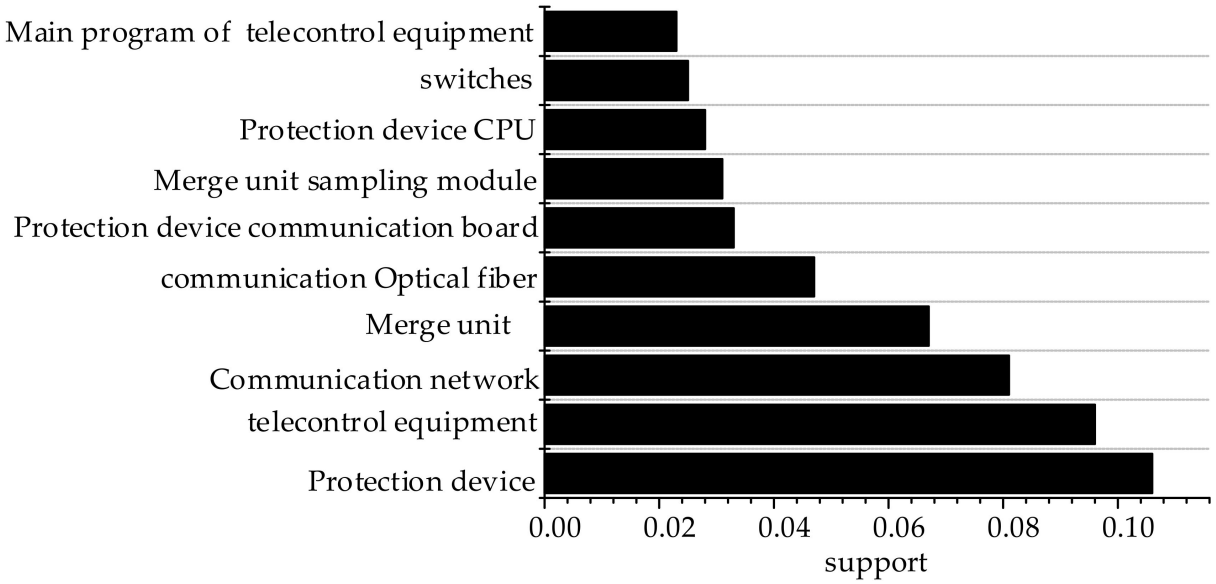


Figure 5. Support of each device.

As can be seen from Figure 5, the equipment with high defect frequency includes protection device, telecontrol equipment, merging unit, secondary loop communication network, and switch. In terms of the specific fault location of the equipment, the defect probability of the communication fiber is the highest, followed by the communication board of the protection device, the sampling module of the merging unit, the CPU of the protection device, and the main program of the remote device.

The above frequent itemsets are all mined by the H-mine algorithm. To illustrate the superiority of the H-mine algorithm, this paper compares the running speed of the Apriori algorithm, FP-growth algorithm, and H-mine algorithm when mining alarm signals, and the support threshold is set as 0.01. Their running time is shown in Figure 6.

It can be seen that the running speed of the H-mine algorithm is better than that of the traditional Apriori algorithm and FP-growth algorithm. Moreover, with the increase of the amount of data, its superiority becomes more obvious.

5.2. Analysis of Mining Association Rule Results

This paper focuses on mining association rules, which can reflect the hidden relationship between transactions. By mining association rules in defect records of smart substation secondary system, the hidden information in defect records can be found to facilitate maintenance personnel to make more reasonable decisions.

This paper analyzes the defect data of secondary equipment as follows:

1. This paper analyzes the association rules between the manufacturer and the faulty equipment to find familial defects.

2. This paper analyzes the association rules between alarm signals and fault causes to reference maintenance personnel.
3. This paper analyzes the association rules between the faulty equipment and the specific fault parts of the equipment to facilitate the maintenance personnel to repair the weak parts of the equipment.

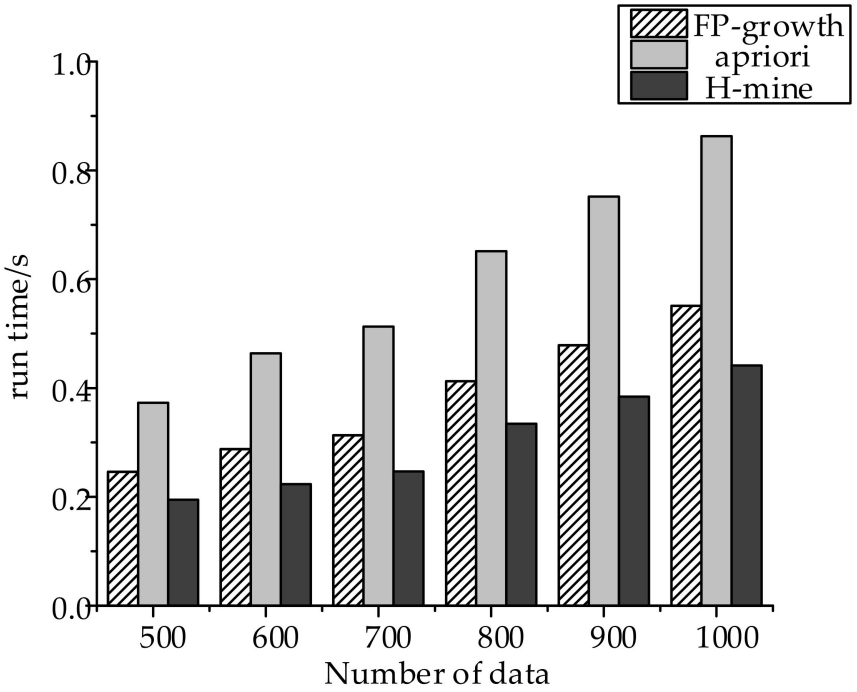


Figure 6. Algorithm performance comparison chart.

The main purpose of the association analysis algorithm is to obtain the corresponding relationship, such as A→B, that meets the user's requirements. A and B are the prefix and suffix, where:

1. When analyzing the relationship between the manufacturer and the cause of the fault, A is the manufacturer, and B is the cause of the fault.
2. When analyzing the cause of the fault, A is the alarm signal, and B is the cause of the fault.
3. When looking for the relationship between the equipment and the fault location, A is the name of the equipment, and B is the fault location.

In view of the above three situations, this paper uses the AFWA algorithm to optimize the key threshold parameters of association analysis and then combines the H-Mine algorithm to realize rapid mining of frequent item sets. Finally, based on the confidence threshold and *lift* threshold obtained by the AFWA algorithm, frequent itemsets are used to generate association rules whose confidence and *lift* are both greater than the threshold.

The operating parameters of the AFWA algorithm are as follows: the population size is 8, the number of fireworks ranges from 2 to 50., the number of Gaussian sparks is 10, the search range of support threshold and confidence threshold is (0,1), the search range of *lift* threshold is (1,10), and the maximum explosion radius is the maximum variation range of fireworks in each dimension.

Some association rules generated under the three conditions are shown in Table 3. The confidence of each rule is as in Figure 7.

Table 3. Association rules.

Number	The Prefix	The Suffix
1	Manufacturer C	Protection device fault
2	Manufacturer A	Merging Unit fault
3	Manufacturer B	combined unit fault
4	Protection device SV chain breaking, merging unit GOOSE chain breaking, merging unit SV chain breaking, measurement and control device SV chain breaking, protection device locking	Merge unit communication board fault
5	Communication interruption of 110 kV protection device opening, interruption of smart terminal GOOSE, abnormal longitudinal channel of protection device, exit of the longitudinal channel	The longitudinal channel of the protection device fault
6	Intelligent terminal GOOSE broken chain, power loss alarm, protection device GOOSE broken chain, measurement and control device GOOSE broken chain	Power fault of intelligent terminal
7	Protection device SV, Protection device GOOSE broken chain, intelligent terminal GOOSE broken chain, protection device lock, reclosing lock	Line protection device communication board fault
8	Manufacturer C, protection device	protection device communication board
9	Manufacturer E, Protection device	pilot protection channel
10	Manufacturer B, Fault Recorder	Communication transmission device

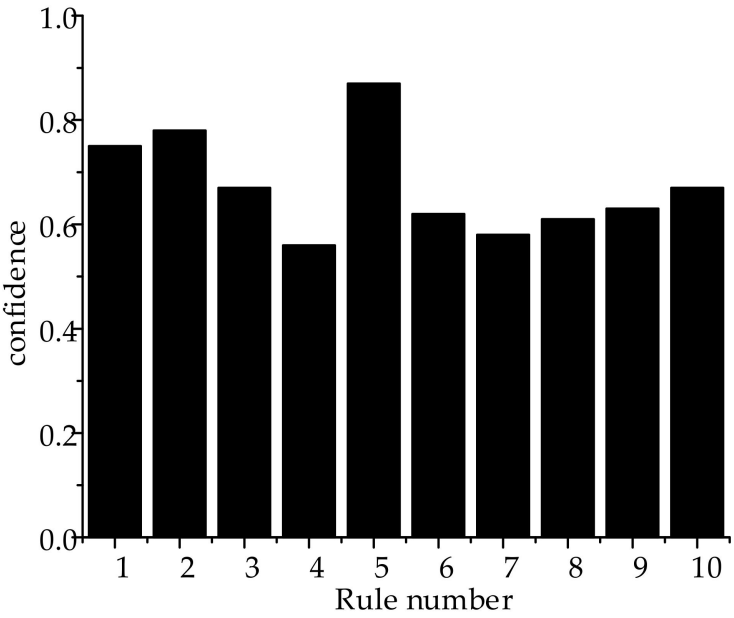


Figure 7. Association rules' confidence.

In particular, when analyzing the cause of the fault, the association rules will only be retained in the simplified form after association rule screening, which not only improves the reference value of the rules but also improves their comprehensibility. The similarity threshold in this paper is set to 0.7.

Rule 4–7 can be simplified as in Table 4. The confidence of rule 4–7 after processing is as in Figure 8.

Table 4. Association rules after processing.

Number	The Prefix	The Suffix
4	Protection device SV broken chain, merging unit GOOSE broken chain, merging unit SV broken chain, measurement and control device SV broken chain	Merge unit communication board fault
5	110 kV protection device open communication interruption, abnormal longitudinal connection channel of protection device, longitudinal connection channel exit	Protection device longitudinal connection fault
6	Intelligent terminal GOOSE broken chain, power loss alarm, intelligent terminal SV broken chain	Power fault of intelligent terminal
7	Protection device SV, GOOSE broken chain, intelligent terminal GOOSE broken chain, protection device locking	Protection device communication board fault

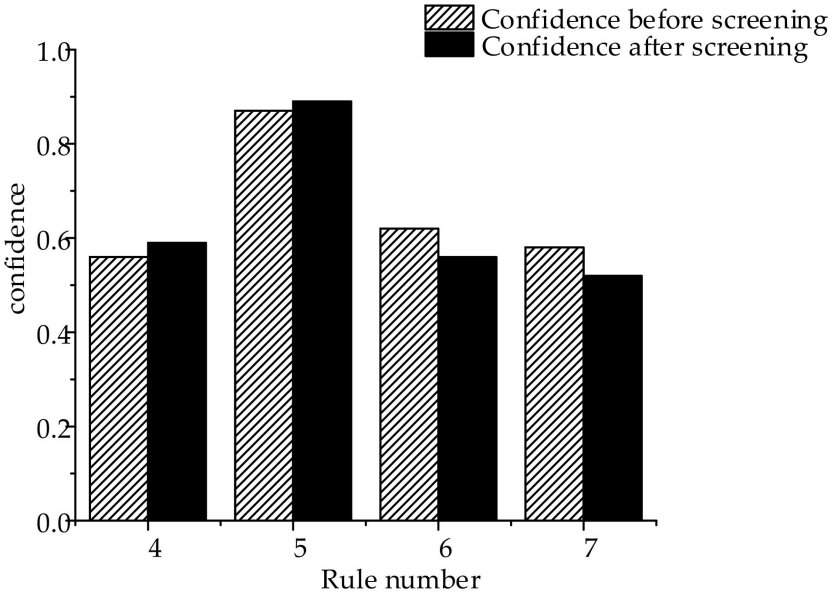


Figure 8. Confidence of the processed association rules.

According to rules 1–3, the switch produced by manufacturer A is more likely to fail, and the confidence is 78%. Similarly, it can be seen that among the equipment produced by manufacturer B and manufacturer C, the equipment that is more prone to fault is the merging unit and protection device, with a confidence of 75% and 67%, respectively. Therefore, these rules can provide the equipment acceptance personnel with reference to improve the acceptance standard when receiving the equipment.

According to the processed rule 4, when the alarm signal is the communication interruption of the 110 kV protection device, the longitudinal channel of the protection device is abnormal, and the longitudinal channel exits, the alarm signal has a strong correlation with the fault cause of the longitudinal channel fault of the protection device, and the confidence is 75%. Similarly, the prefixes and suffixes of Rule 5–7 represent that alarm signals are strongly correlated with their corresponding fault causes, and their confidence is greater than 50%. By comparing Figure 4 with Figure 3, it can be found that although the confidence of some association rules decreases, the number of prefixes of rules decreases; that is, the number of alarm signals decreases. In this way, some derivative signals can be filtered out, and the comprehensibility of rules can be improved. This kind of rule can help maintenance personnel to determine the probable cause of the fault when there is an alarm signal and provide a basis for subsequent repair work.

As can be seen from Rule 8–10, when the conditions are manufacturer B and the fault recorder, the conclusion is that the defective part is the communication transmission device, and the confidence is 67%. Therefore, it can be seen that the weak link of the fault recorder of manufacturer B is likely to be in the communication transmission device. Therefore, this kind of association rule can provide the maintenance personnel with weak links in secondary equipment and help the maintenance personnel to make the maintenance plan.

To verify the searching ability of the AFWA algorithm, this paper uses 10-fold cross-validation to compare the optimization effect of the AFWA algorithm, FWA algorithm, and SPSO algorithm in association analysis.

The weight of the fitness function in this paper is set as W1 = 0.8 and W2 = 0.2. The running parameters of FWA algorithm are the same as those of AFWA. The operating parameters of SPSO are as follows: the maximum value of inertia weight is 0.8, the minimum value is 0.1, and the learning factors of C1 and C2 are both 2.

To evaluate the quality of the association rule set generated by the algorithm, the following indexes are set for evaluation: average *lift* and *CF*. Moreover, to make the comparison process more scientific, the number of iterations set in this paper is all 300 times. Figure 9 is the comparison of the three algorithms in the three cases.

According to Figure 9a,b, in the first case, the optimization effect of the three algorithms is the same in Experiment 1 and Experiment 6. However, in Experiment 2, Experiment 4, Experiment 5, Experiment 7, Experiment 9, and Experiment 10, the association rule *lift* and *CF* optimized by the AFWA algorithm are all the highest. In Experiment 3, the quality of association rules optimized by the AFWA algorithm was slightly lower than that of the traditional AFWA algorithm but higher than that of the SPSO algorithm. There were eight kinds of experiments, and the index of association rules optimized by the AFWA algorithm was the lowest among the three algorithms.

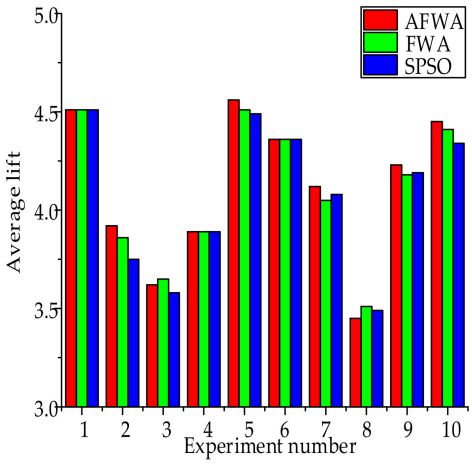
According to Figure 9c,d, in the second case, in Experiment 1, 3, 4, 5, 6, 7, and 9, the *lift* and *CF* of association rules optimized by the AFWA algorithm are the highest. In Experiment 8 and Experiment 10, The quality of association rules optimized by the AFWA algorithm is lower than that optimized by the SPSO algorithm but higher than that of the FWA algorithm. In Experiment 2, the quality of association rules optimized by the AFWA algorithm is lower than that of the FWA algorithm but better than that of the SPSO algorithm.

According to Figure 9e,f, in the third case, the optimization effects of the three algorithms are the same in Experiment 2. In Experiment 4, the *lift* and *CF* of association rules optimized by AFWA algorithm are slightly lower than FWA algorithm and higher than the SPSO algorithm. In other cases, the *lift* and *CF* of association rules optimized by the AFWA algorithm are the highest.

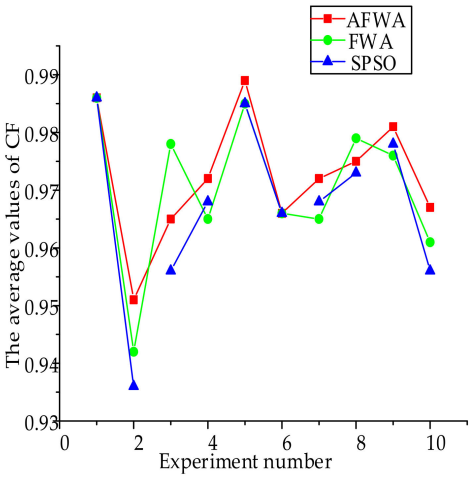
The fitness ranking of the three algorithms in the three situations is as in Figures 10–12.

In the first case, the average ranking of the three algorithms is 1.36, 2.07, and 2.64; in the second case, the average ranking of the three algorithms is 1.3, 2.45, and 2.25; in the third case, the average ranking of the three algorithms is 1.24, 2.33, and 2.33. Therefore, it can be seen that the overall ranking of the AFWA algorithm is better than the other two algorithms when searching the threshold parameters of the association analysis for the defect data of the secondary system.

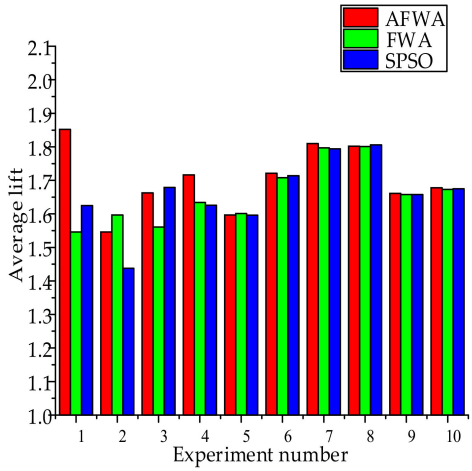
In addition, to illustrate the advantages of the proposed method, this paper compares the differences between the two methods by comparing the traditional association analysis.



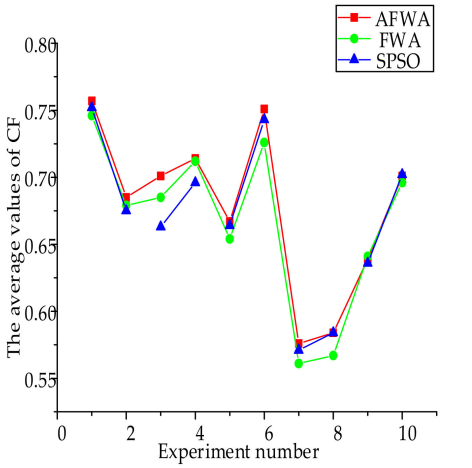
(**a**) Average *lift* of association rules in the first case.



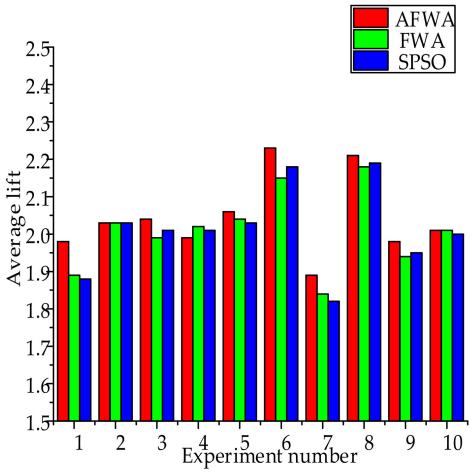
(**b**) Average *CF* of association rules in the first case.



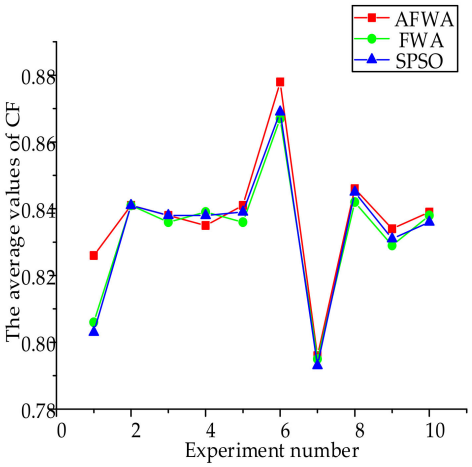
(**c**) Average *lift* of the association rules in the second case.



(**d**) Average *CF* of the association rules in the second case.



(**e**) Average *lift* of association rules in the third case.



(**f**) Average *CF* of association rules in the third case.

Figure 9. Comparison diagram of the algorithm results.

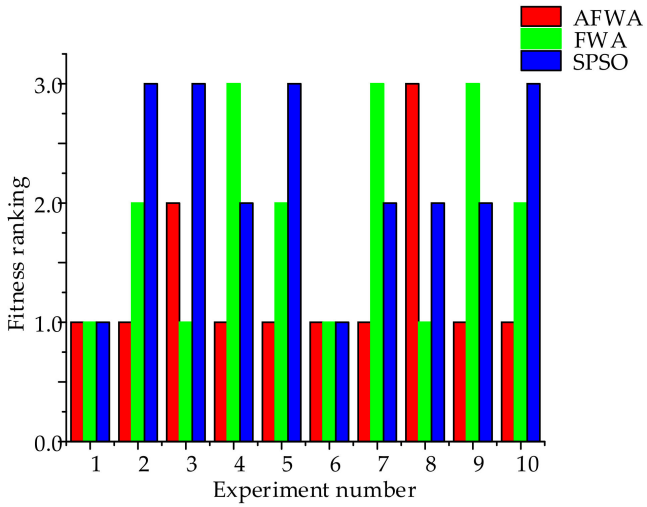


Figure 10. Algorithm fitness ranking of the first case.

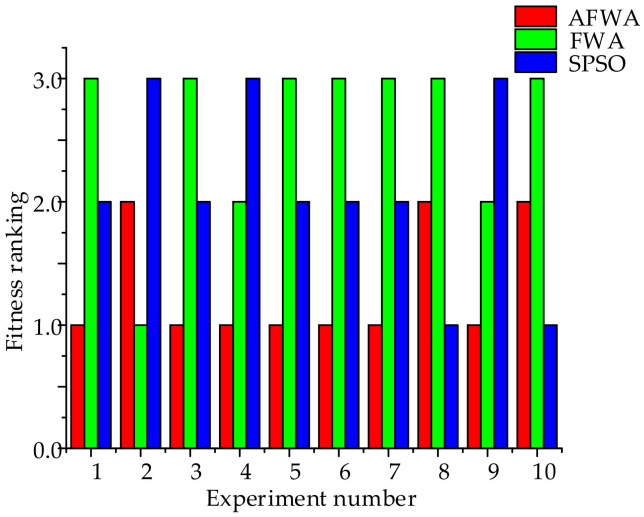


Figure 11. Algorithm fitness ranking of the second case.

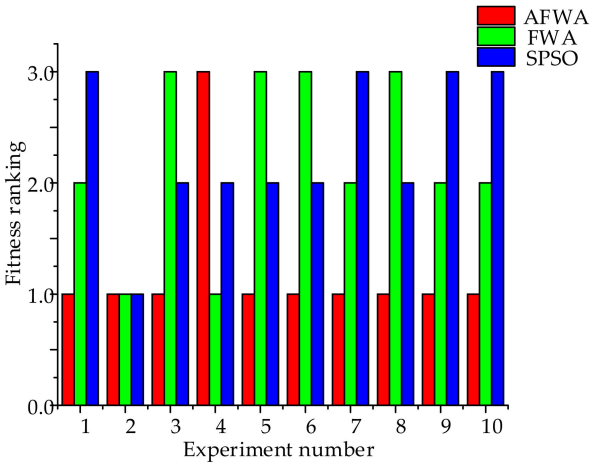


Figure 12. Algorithm fitness ranking of the third case.

In this paper, the average similarity is set to illustrate the benefit of this method compared with traditional association analysis, which is defined as the average of the similarity among all rules. The calculation formula is as follows:

$$AveSimi = \frac{\sum_{i=1}^{Num-1} \sum_{j=2}^{Num} (Simi[i,j] + Simi[j,i])}{(Num \times (Num-1))} \quad (24)$$

Since most traditional association analysis algorithms have similar principles, the basic Apriori algorithm is chosen as the comparison algorithm in this paper. The traditional Apriori algorithm only has the concepts of support and confidence. In this paper, the support and confidence are set to be consistent with the optimization results of AFWA to compare the advantages and disadvantages of the two key parameters in three situations. Figures 13–15 are the differences of *lift*, *CF*, and similarity in the three cases, where horizontal axis 1–1 represents the AFWA-H-Mine algorithm in the first case, 1–2 represents the Apriori algorithm, and so on.

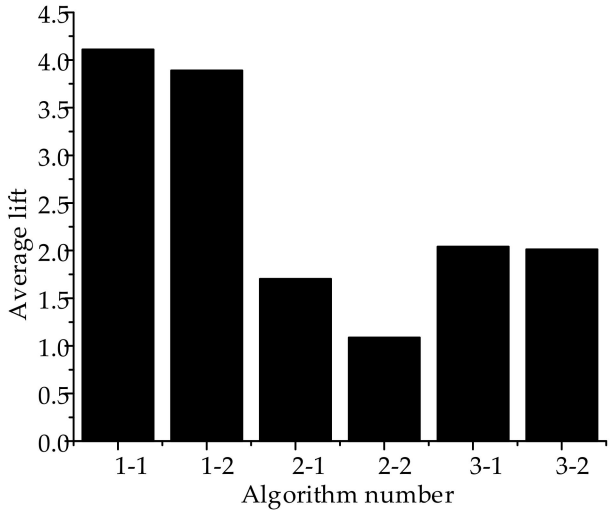


Figure 13. Comparison between AFWA-H-Mine and traditional algorithms of the first case.

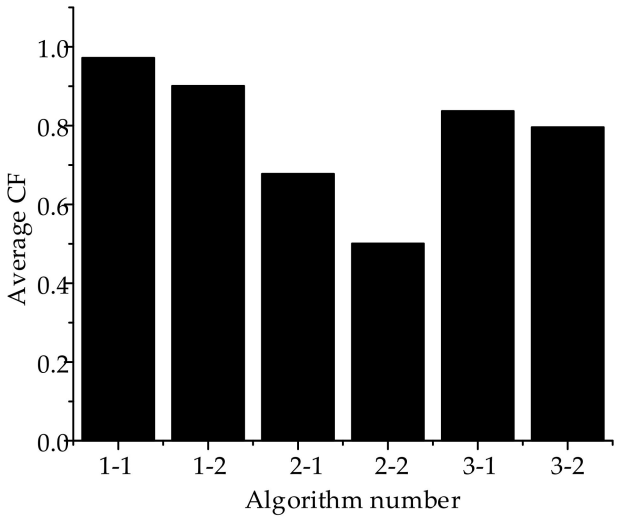


Figure 14. Comparison between AFWA-H-Mine and traditional algorithms of the second case.

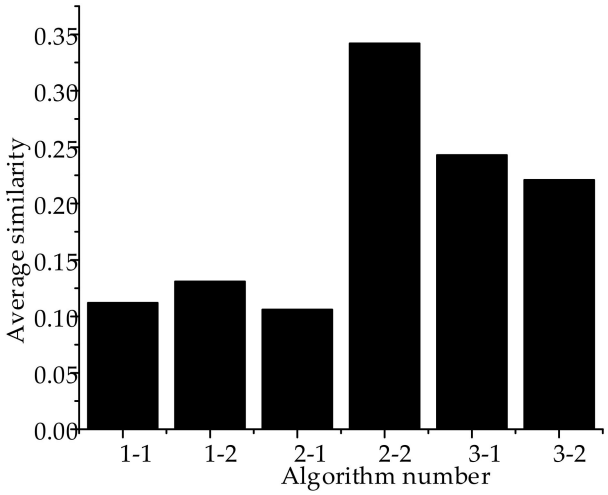


Figure 15. Comparison between AFWA-H-Mine and traditional algorithms of the third case.

As can be seen from Figures 13–15, the AFWA-H-Mine association rule optimization method proposed in this paper has a high quality of rule number generated by AFWA-H-Mine in all three cases, and its *lift* and *CF* are superior to the traditional Apriori algorithm. When analyzing the cause of fault, because the number of rules generated is the largest, most rules are to be eliminated, so the optimization effect is the best. The number of rules is small in the other two cases, so the optimization effect is not obvious. In addition, the introduction of the rules filtering strategy reduces the rule length during fault analysis, with the average rule length reduced from 6.01 to 4.09.

6. Conclusions

Aiming at the hidden association rules in the defect record of the secondary equipment of smart substation, this paper classifies three situations, which are: the manufacturer and the faulty equipment, the alarm signal and the cause of the fault, and the specific fault position of the faulty equipment and equipment. The association analysis algorithm is used to mine the above three cases, respectively, and the association rules are obtained to provide auxiliary suggestions for maintenance personnel.

This paper proposes a defect data association analysis model based on the H-Mine algorithm. At the same time, the AFWA algorithm is used to optimize the key threshold. Compared with the traditional fireworks algorithm and SPSO algorithm, the AFWA algorithm can set different search resources for individuals with different fitness in the population and can adjust the step size adaptively. The analysis of an example shows that the search performance of the algorithm is better.

Compared with the traditional association analysis model, this model uses the fitness function, composed of the average *lift*, average *CF*, and data coverage rate, without manually adjusting the threshold value. By comparing the average *lift* and average *CF* of the association rules generated by these two methods, the results show that the quality of the association rules generated by the AFWA-H-Mine algorithm is higher.

Author Contributions: Conceptualization, M.W. and Y.X.; methodology, M.W.; software, M.W.; validation, Y.X. and W.F.; formal analysis, M.W., Y.X. and W.F.; investigation, M.W.; resources, Y.X.; data curation, W.F.; writing—original draft preparation, M.W.; writing—review and editing, M.W. and W.F.; visualization, Y.X.; supervision, Y.X.; project administration, Y.X.; funding acquisition, Y.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by "National Key Research & Development Program of China, grant number 2016YFB0900203".

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chen, K.; Mahfoud, R.J.; Sun, Y.; Nan, D.; Wang, K.; Alhelou, H.; Siano, P. Defect texts mining of secondary device in smart substation with GloVe and attention-based bidirectional LSTM. *Energies* **2020**, *13*, 4522. [CrossRef]
2. Zhang, Y.; Wang, H.; Diao, X.; Tong, X.; Guo, S.; He, X. Integrated risk assessment of intelligent substation secondary system considering the protection failure. *Power Syst. Prot. Control* **2018**, *46*, 155–163.
3. Zhang, Y.; Hu, C.; Huang, S.; Feng, S.; Lin, G. Apriori algorithm based data mining and analysis method for secondary device defects. *Autom. Electr. Power Syst.* **2017**, *41*, 147–151.
4. Dai, Z.; Lu, H.; Liu, Y.; Liu, B.; Chen, Y. A fault diagnosis method for the secondary circuits of protection systems in smart substations based on improved D-S evidence theory. *Power Syst. Prot. Control* **2020**, *48*, 59–67.
5. Fang, X.; Huang, W.; Ye, D.; Huang, Y. Application of a distributed parallel FP-growth algorithm in secondary device defects monitoring. *Power Syst. Prot. Control* **2021**, *49*, 160–167.
6. Fedushko, S.; Ustyianovych, T.; Gregus, M. Real-time high-load infrastructure transaction status output prediction using operational intelligence and big data technologies. *Electronics* **2020**, *9*, 668. [CrossRef]
7. Ren, B.; Zheng, Y.; Wang, Y.; Sheng, S.; Li, J.; Zhang, H.; Zheng, C. Research on fault location of secondary equipment in smart substation based on deep learning. *Power Syst. Technol.* **2021**, *45*, 713–721.
8. Zhang, X.; Chen, Q.; Sun, M.; Huang, W.; Wang, L.; Liu, B. Fault tracking of high-voltage circuit breakers in case of secondary circuit faults in intelligent substations. *Electr. Power Autom. Equip.* **2020**, *40*, 212–217.
9. Tian, M.; Zhang, L.; Guo, P.; Zhang, H.; Chen, Q.; Li, Y.; Xue, A. Data dependence analysis for defects data of relay protection devices based on apriori algorithm. *IEEE Access* **2020**, *8*, 120647–120653. [CrossRef]
10. Chen, Y.; Li, S.; Zhang, L.; Lu, H.; Dai, Z. Association analysis for defect data of secondary device in smart substations based on improved Apriori algorithm. *Power Syst. Prot. Control* **2019**, *47*, 135–141.
11. Xiao, Y.; Liu, S.; Jian, W.; Song, J. A kind of defects analysis method for secondary device of substation based on fp-growth algorithm. *Electr. Meas. Instrum.* **2020**, *57*, 83–90.
12. Pei, J.; Han, J.; Lu, H.; Nishio, S.; Tang, S.; Yang, D. H-Mine: Fast and space-preserving frequent pattern mining in large databases. *IIE Trans.* **2007**, *39*, 593–605. [CrossRef]
13. Sarath, K.; Ravi, V. Association rule mining using binary particle swarm optimization. *Eng. Appl. Artif. Intell.* **2013**, *8*, 1832–1840. [CrossRef]
14. Kuo, R.; Chao, C.; Chiu, Y. Application of particle swarm optimization to association rule mining. *Appl. Soft Comput.* **2011**, *11*, 326–336. [CrossRef]
15. Dong, D.; Ye, Z.; Cao, Y.; Xie, S.; Wang, F.; Ming, W. An improved association rule mining algorithm based on ant lion optimizer algorithm and FP-growth. In Proceedings of the 2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, Metz, France, 18–21 September 2019; pp. 458–463.
16. Tan, Y.; Zhu, Y. Fireworks algorithm for optimization. In Proceedings of the International Conference in Swarm Intelligence, Beijing, China, 12–15 June 2010; Springer: Berlin/Heidelberg, Germany, 2010; pp. 355–364.
17. Zheng, S.; Janecek, A.; Tan, Y. Enhanced fireworks algorithm. In Proceedings of the 2013 IEEE Congress on Evolutionary Computation, Cancun, Mexico, 20–23 June 2013; pp. 2069–2077.
18. Li, J.; Zheng, S.; Tan, Y. Adaptive fireworks algorithm. In Proceedings of the 2014 IEEE Congress on Evolutionary Computation, Beijing, China, 6–11 July 2014; pp. 3214–3221.
19. Kennedy, J.; Eberhart, R. Particle swarm optimization. In Proceedings of the ICNN'95-International Conference on Neural Networks, Perth, WA, Australia, 27 November–1 December 1995; pp. 1942–1948.
20. Zhan, Z.; Zhang, J.; Li, Y. Adaptive particle swarm optimization. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **2009**, *39*, 1362–1381. [CrossRef] [PubMed]