

Research Article

An Algorithm for Mining of Association Rules for the Information Communication Network Alarms Based on Swarm Intelligence

Yang Wang,^{1,2} Guocai Li,^{1,2} Yakun Xu,^{1,2} and Jie Hu^{1,2}

¹ MicroNano System Research Center, College of Information Engineering, Taiyuan University of Technology, Taiyuan, Shanxi 030024, China

² Key Lab of Advanced Transducers and Intelligent Control System of the Ministry of Education, Taiyuan University of Technology, Taiyuan, Shanxi 030024, China

Correspondence should be addressed to Jie Hu; hujie0351@gmail.com

Received 5 August 2013; Accepted 17 November 2013; Published 19 January 2014

Academic Editor: Orwa Jaber Housheya

Copyright © 2014 Yang Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Due to the centralized management of information communication network, the network operator have to face these pressures, which come from the increasing network alarms and maintenance efficiency. The effective analysis on mining of the network alarm association rules is achieved by incorporating classic data association mining algorithm and swarm intelligence optimization algorithm. From the related concept of the information communication network, the paper analyzes the data characteristics and association logic of the network alarms. Besides, the alarm data are preprocessed and the main standardization information fields are screened. The APPSO algorithm is proposed on the basis of combining the evaluation method for support and confidence coefficient in the Apriori (AP) algorithm as well as the particle swarm optimization (PSO) algorithm. By establishing a sparse linked list, the algorithm is able to calculate the particle support thus further improving the performance of the APPSO algorithm. Based on the test for the network alarm data, it is discovered that rational setting of the particle swarm scale and number of iterations of the APPSO algorithm can be used to mine the vast majority and even all of the association rules and the mining efficiency is significantly improved, compared with Apriori algorithm.

1. Introduction

The operation and maintenance management of information communication network mainly refers to timely discovery, locating and handling of any network fault to ensure smooth and efficient operation as well as guarantee in major emergencies pertinent to network operation, complaints about network quality from customers, assessment and analysis of network quality, prediction of planning, construction, and so forth. The time consumed during fault location and judgment in the application layer of a large-scale network accounts for 93% of its total time for failure of recovery [1]. The huge network structure and multifunctional device types also bring about large amounts of alarm data due to such characteristics of the information communication network

as topological structure densification, network device micro-miniaturization, communication board precision, and so forth. Therefore, the foundation of the network operation and maintenance is the effective management of the network alarms.

As an important supporting means for network operation and maintenance management, network management system directly influences the quality of service which the information communication network provides to its customers [2]. The network management system is developing toward integrated service network management update from independent device network management, manufacturer device network management, and integrated professional network management. The centralized monitoring management function of the professional information communication network

operation management will make problems exhibit a sharp full data increasing, including network faults, device alarms, and customer complaints.

As the information communication system consists of various medium interlinked network devices and operating systems implicit and complex-correlated logic is ubiquitous among network elements; that is, a certain fault point may trigger numerous alarms in the whole network. The sudden intensive alarms not only consume the resources of the network management system but also obscure the position of the network fault source points thus severely impeding trouble shooting by the network operation and maintenance personnel. Several alarms are incorporated into a single alarm or source alarm with a large amount of information by such links as paraphrasing and explaining, eliminating and filtering, information integration, and correlating and transforming, and so forth. It aims at assisting the operation and maintenance personnel to analyse fault messages and locate faults quickly, that is, mining analysis on alarm association rules.

Mining of alarm association rules refers to a process of analysis on the association between the attributive characteristic logic of the alarms within devices and the topological hierarchy of network devices. It aims at achieving clear critical alarms, accurate fault location and troubleshooting, and intelligent fault prediction and evaluation. The mining of alarm association rules can be divided into three levels: analysis on alarm association in the device within the profession, analysis on topological alarm association of the network device within the profession, and analysis on inter-professional topological alarm association of the network device but their core is mining algorithm for association rules [3].

The centralized management of information communication network brings about large amounts of alarm data. A rapid mining analysis on the network alarm association rules is achieved by the classic Apriori association mining algorithm and PSO algorithm under the context of big data. The alarm association relationship can be used to add and merge the fault alarms, maintain the work order, improve the centralized monitoring efficiency, and reduce the cost of network maintenance.

The Apriori is an association rules mining algorithm based on characteristics of frequent item sets (priori knowledge) whose core concept is a layer-wise iterative search of the theory of frequent item sets. However, the Apriori algorithm thought also presents some inevitable problems. For instance, frequent repeated scans of the information in the sample database lead to a heavy load on the system I/O; large item sets lead to a sharp increase in the number of the candidate frequent item sets and a significant increase in operation time, and so forth.

Swarm intelligence refers to the macroscopic intelligent group behavior showed by various types of organism individuals in the nature during survival, collaboration, and evolution. Application research is conducted for the swarm intelligence algorithm in optimization solutions of engineering problems such as economic analysis and forecast, structural damage positioning and inspecting, command and

dispatch of communication and transportation, evacuation route planning, target identifying and tracking, factory site selection and evaluation, communication network planning, and route plan preparation [4]. The swarm intelligence algorithm has such advantages as distributed control, indirect information transfer, and simple individuals and swarm intelligence. As a classic swarm intelligence algorithm, the particle swarm optimization also has the above characteristics.

Centralized management of the alarms in the information communication network is an important part of operation maintenance of the information communication network. The alarm correlation directly influences the quantity and quality of the alarm work orders. An analysis on the large amounts of alarm data through an efficient algorithm becomes the critical technical means. The APPSO discussed in the paper incorporates the Apriori algorithm and swarm optimization algorithms and applies swarm optimization algorithms in the information communication field.

Section 2 in the paper elaborates such basic concepts of faults in the information communication network, network alarms, alarm standardization, and so forth; Section 3 discusses the data characteristics of network alarms, and the alarm correlation logical relationships within and between network devices; Section 4 describes achieving quality improvement of the data source of the network alarms by pre-processing of the network alarm data; Section 5.1 presents the concepts of support and confidence coefficient and mining analysis process in Apriori algorithm in combination with examples; Section 5.2 describes the swarm intelligence model and basic flow of the PSO algorithm; Section 5.3 discusses the creation of APPSO association rule mining algorithm, which deducts on the basis of the Apriori and PSO algorithm characteristics. Besides, combining with the characteristics of the network alarm data, the section puts forward the improvement of the performance of the APPSO association rule mining algorithm by sequencing code, sliding window, sparse linked list, and nature of the Apriori algorithm. It conducts a performance test for the algorithm through the alarm data in the information communication network from different angles. At the end of the section, an index evaluation of the alarm association rate is put forward, which is used for application of the alarm correlation relationship derived from the APPSO algorithm mining into the actual network.

2. Concepts Pertinent to Alarms in the Information Communication Network

Concepts pertinent to the data analysis on the alarms in the information communication network are defined as follows [5, 6].

Definition 1. A network fault refers to an event where the information communication network is not able to operate normally and efficiently due to some reasons and even no service can be provided. The reasons causing network faults can be divided into network device faults, communication link abnormality, inappropriate operation and maintenance,

energy power and room environment abnormality, and network system faults (affecting monitoring instead of the communication service).

Definition 2. A network alarm is a message triggered during abnormal operation of communication device and each alarm message represents its unique running status. No uniform standard specification is applicable to the network devices in the whole industry due to the difference in mechanism and connotation of the alarm messages of devices of different types from various manufacturers. However, the standardization can be achieved by specific standardized fields.

Definition 3. Alarm standardization redefines the level, classification, influence, and so forth, of the full professional alarms, which achieving the target on achieve mapping definition, normative classification, and centralized management of professional alarms of different manufacturers.

Definition 4. The alarm standardization fields include profession, manufacturer, device type, alarm title, auxiliary fields of alarm explanation, manufacturer alarm level, applicable manufacturer version number, network management alarm level, network management alarm ID, alarm explanation, alarm class, alarm logic class, alarm logic subclass, effect of such an event on the device, effect of such an event on the service, and standard name of the alarm.

Definition 5. The alarm standardization fields of the network management system refer to the other alarm standardization fields of the network management system excluding the alarm standardization fields, for example, city/county/district, network element name, number of network element board card, local port information of the alarm, remote port information of the alarm, occurrence time of the network element alarm, discovery time of the network management alarm, elimination time of the alarm, and so forth.

3. Data Characteristics and Association Logic of Network Alarms

The information communication network has such characteristics as complex, hierarchical, and full end-to-end networking. These network elements have certain physical and logical association, and the independent network element failure will result in “click alarm, multiclick dissemination” effect on related network element. However, there is association of occurrence time and logical name between these alarms. Thus, association, classification, and combination of such alarms can substantially improve the efficiency of centralized monitoring [7].

3.1. Data Characteristics of Network Alarms. Information communication network alarm is characterized by huge data volume, alarm fluctuation, network communication effect, accumulative and lagging effects and redundancy of fault messages, and so forth. The analysis of these characteristics

will contribute to mining analysis on rules of association among alarms.

(1) *Huge Data Volume.* The number of alarms and faults in the current network is huge due to such characteristics as diversification of types of information communication network services, network scale expansion, topological structure tightness and centralization of network monitoring, and so forth.

(2) *Alarm Fluctuation.* From the perspective of monitoring management, the equipment failure alarms have certain unpredictability. The crash of critical equipment will cause the whole network paralysis leading to a sharply increasing number of alarms inevitably. Similarly, the alarms can be eliminated if the failures are maintained and handled timely. For instance, the block of central transmission lines will affect local lines, lines across cities, and relative network equipment; thus, all relevant equipment exhibits alarm conditions. If the central lines are dealt with appropriately, the alarm will be removed rapidly.

(3) *Network Communication Effect.* The alarm does not spread through some concrete networks but relies on the independent “management network” [8]. Take SDH network alarm for example, LAN regenerator section LOS alarm → multiplex section MIS → AIS alarm → remote device MS-FERF alarm connected to local devices and AU-AIS alarm → local HO-VC HP-AIS alarm → local TU-AIS alarm and HP-FERF/RDI alarm.

(4) *Accumulative and Lagging Effect.* The abnormality of some network equipment would degrade the relative network quality. If this condition has accumulated to an extent that exceeds the limits, the connected network equipment would alarm. Besides, these features may be caused by clock synchronous exception among communication equipment, NM for manufacturer’s equipment and NM for multidisciplinary or abnormal network management data.

(5) *Redundancy of Fault Messages.* Fault points on single panel would cause the associated devices parts to alarm; and the failure of network convergence nodes can trigger a large-scale network alarm. For example, the failure of MSC server (mobile switching center) will lead some devices to stay in an alarm state, such as, MGW (media gateway), BSC (base station controller), and RNC (radio network controller). And this phenomenon will lead to a sudden “alarm storm.”

(6) *Abundant Property Field.* Each alarm corresponds to some recognized information combination. Different property fields reveal certain relevant logic.

(7) *Abnormal Alarm.* It can be divided into waste alarm, ultrashort alarm, and overlength alarm. The waste alarm is not caused by the filter clear of network access test and device data in time. The ultrashort alarm points the alarm lasts for less than one minute. And the overlength alarm refers to the alarms which are not removed after a long time.

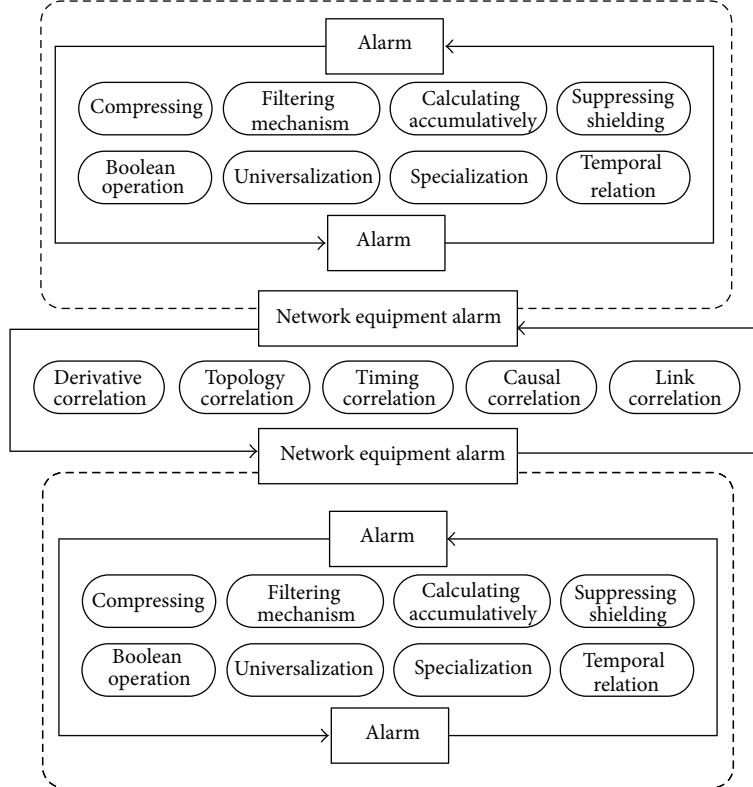


FIGURE 1: Association logic of network alarms.

3.2. Association Logic of Network Alarms. The network association logic can be divided into two levels, that is, alarm association logic within network device and alarm association logic among network device as shown in Figure 1.

The alarm logical association on the network equipment itself is as follows [9]: (1) alarm compressing: taking the simultaneous multialarm which has the same attributes (adjacent cells, same network element or light path, etc.) into an alarm; (2) filtering mechanism: alarm which does not conform to the attribute association will be deleted; (3) calculating accumulatively: a number of concurrent alarms will be converted to an alarm with new name; (4) suppressing shielding: low priority alarms will be suppressed when they are of high priority to be generated; (5) boolean operation: making a group of alarms in conformity with some rules of Boolean operation into an alarm; (6) generalization: network element is to be a more general alarm; (7) specialization: the more detailed alarm information will replace network element alarms; (8) temporal relation: the different alarms are to be generated as per certain time sequence.

Alarm association among groups of network equipment is as follows: (1) derivative association: the network equipment alarms are divided into root alarm and derivative alarm; (2) topological association: the network equipment alarm contains home terminal alarm and opposite end; (3) timing association: the same fault point generates alarms with the same time trigger characteristic; (4) causal association: Occurrence of Alarm A causes Alarm B, that is, element management system has been out of management as a result of optical cable break; (5) link association: convergence line

fault will trigger the network equipment alarm on the entire path and send unification orders.

4. Preprocessing of Network Alarm Data

The transmission network device alarm data is used as the analytical data for association rules for the information communication network alarms and the link of data preprocessing is as follows (Figure 2).

(1) *Data Extraction.* All transmission alarms within a specific time interval are extracted through the network management system (including engineering cutover and device alarms arising from network adjustment) and the data fields extracted include alarm standardization field and network system alarm standardization field.

(2) *Data Cleaning.* Special data affecting the algorithm analysis quality is cleaned from the alarm data extracted and such data includes

- ① abnormal data: junk alarm, ultrashort alarm, ultra-long alarm, and abnormal and special alarm data,
- ② incomplete data: alarm data with a null alarm determinant attribute field,
- ③ erroneous data: alarm data with a large difference between the time field of the network management alarm and the time field of the device alarm due to time synchronization abnormality,

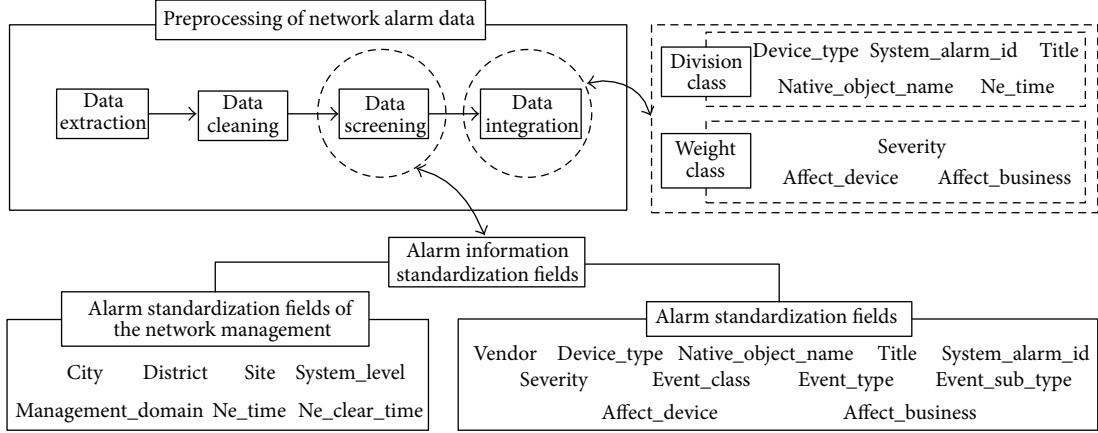


FIGURE 2: Preprocessing of network alarm data.

- ④ duplicated data: duplicated alarm data due to merging or removing flashes.

(3) *Data Screening*. ① Interference data: screen and reject the interference alarm data, for example, uncorrelated alarms (alarms such as access control enabling and mismatching of the main and standby sing board versions) in a number of signal alarms (alarms such as signal degradation indication and output signal loss) are rejected. During screening, the duplicated alarms should not be deleted blindly and they should be analyzed and discriminated based on the actual fault conditions considering that the duplicated alarms may be caused by different faults during different periods [10].

② Alarm information standardization field: main information fields are screened from the standardization fields of the network management alarms and alarm standardization fields for subsequent mining of association rules. These information fields are set as two classes: division class and weight class. The alarm information fields of division class are mainly used to describe attribution relation and attribute parameters of alarms. The alarm information fields of weight class are mainly used to describe importance difference and influence and assign differentiated weight to the data of the association rule mining algorithm.

(4) *Data Integration*. The alarm processed in the above link and its corresponding information standardization fields are resorted out eventually and generate network alarm data sources with high information amount.

5. Mining Algorithm for Association Rules for the Network Alarm Data

The Apriori algorithm has been widely used by researchers as a classic mining algorithm for association rules. While the swarm intelligence algorithm has been studied deeply and applied in various fields due to its characteristics such as distributed control, low communication overhead, simple behavior rule, and strong self-organization. The APPSO

algorithm is exactly an efficient algorithm for it incorporates the above two algorithm thoughts and combines with the data characteristics of the alarms in the information communication network.

5.1. Example Analysis for Apriori Algorithm. On the ICDM (IEEE International Conference on Data Mining) held in December 2006, the top ten classical algorithms were selected from the 18 candidate algorithms after three links of nomination, review, and voting, that is, C4.5 (classification), K-Meams (statistical learning), SVM (statistical learning), Apriori (association analysis), EM (statistical learning), PageRank (link mining), AdaBoost (bagging and boosting), kNN (classification), Naive Bayes (classification), and CART (classification). The Apriori algorithm formulated by Wu and Vipin in 2009 ranks fourth among the ten top classical algorithms for data mining, which also sufficiently shows its importance in data mining algorithm [11].

The association rules mining algorithm exactly obtains the association relationship among terms from data sets through mathematical logic. The market basket analysis sufficiently embodies the industrial application value of the association rules mining algorithm. The Apriori is an association rules mining algorithm based on characteristics of frequent item sets (a priori knowledge) whose core concept is a layer-wise iterative search of the theory of frequent item sets.

In combination with the examples of fault alarms of the information communication network, the application of concept and flow of the Apriori algorithm are discussed as follows.

5.1.1. Concept of the Apriori Algorithm

- (1) All item sets: all alarm item sets of the examples, that is, Alarm1–Alarm5,
- (2) item set: concurrent item combination, for example, {Alarm1, Alarm2}, {Alarm2, Alarm3, Alarm4},

- (3) support: describes universality and frequency of association rules and the association rule of high support reflects that it may be applicable to most events of the data sets,
- (4) support count: the number of alarm affairs contained in a group of item sets,
- (5) confidence: describes reliability and accuracy of the association rules, that is, probability of Alarm2 occurrence on the premise of Alarm1 occurrence (conditional probability).

As for the mining association rules of the Apriori algorithm, high support and low confidence of the association rule indicate the reliability of the association rule is poor; low support and high confidence of the association rule indicate the applicability of the association rule is poor. Minimum support count and minimum confidence are set manually by users. An association rule is deemed to be concerned if it satisfies both parameters above [12]. The matching relation between the support and the confidence should be set rationally in combination with the value demand for industrial rules in practical application.

The generation process of association rules is also the process where joining, pruning, and enumerating are performed through support and confidence. The association rules are not able to be applied directly through the algorithm; besides, the application value requires analyzing and screening by experts.

5.1.2. Flow of the Apriori Algorithm. The flow of the Apriori algorithm can be reduced to the following steps [13]: (1) analysing the frequent item sets, that is, obtaining all item sets no less than the preset minimum support count from the iteration of the full database (joining, pruning, and enumerating); (2) obtaining the strong association rules, that is, extracting the association rules from the frequent item sets based on the minimum support and minimum confidence. In combination with instances, the analysis and explanation are presented in Table 1.

Table 1 shows the corresponding alarm items generated on the network device when the information network fails. The network fault events are successively defined as Fault 1–Fault 5. The alarm item class corresponding to each fault is defined as Alarm1–Alarm5 (reduced to A1–A5). The network faults arising from different reasons will generate different combinations of alarm item classes (Table 1).

(1) All alarm item sets are scanned and the support of each alarm item is calculated in Table 2.

(2) The minimum support count is 2 and the candidate item set C1 will form after screening (eliminating A5) of the alarm item combinations in L1 (see Table 3).

(3) All alarm item sets are scanned again to form the support calculation L2 based on the candidate item set C1 (see Table 4).

(4) The minimum support count is 2 and the candidate item set C2 will form after screening (eliminating {A1, A4} and {A3, A4}) of the alarm item combinations in L2 (see Table 5).

TABLE 1: Fault alarms of information communication network.

Network fault event (fault ID)	Alarm item sets (alarm items)
Fault 1	A1, A2, A3, A5
Fault 2	A2, A4
Fault 3	A2, A3
Fault 4	A1, A2, A3, A4
Fault 5	A1, A3

TABLE 2: Calculation of support of all alarm item sets L1.

Alarm item	Support count
A1	3
A2	4
A3	4
A4	2
A5	1

TABLE 3: Support of alarm item sets C1.

Alarm item	Support count
A1	3
A2	4
A3	4
A4	2

TABLE 4: Support of alarm item sets L2.

Alarm item	Support count
{A1, A2}	2
{A1, A3}	3
{A1, A4}	1
{A2, A3}	3
{A2, A4}	2
{A3, A4}	1

TABLE 5: Support of Alarm Item Sets C2.

Alarm item	Support count
{A1, A2}	2
{A1, A3}	3
{A2, A3}	3
{A2, A4}	2

TABLE 6: Support of alarm item sets L3.

Alarm item	Support count
{A1, A2, A3}	2
{A1, A2, A4} [*]	1
{A1, A3, A4} [*]	1
{A2, A3, A4} [*]	1

(5) All alarm item sets are scanned again to form the support calculation L3 based on the candidate item set C2 (see Table 6). Based on the nature of the Apriori algorithm (all subsets of the item sets are frequent necessarily), {A1, A4}

and $\{A3, A4\}$ are not frequent item sets. Thus, $\{A1, A2, A4\}^*$, $\{A1, A3, A4\}^*$, and $\{A2, A3, A4\}^*$ in Table 6 are not frequent item sets and can be excluded directly.

(6) The minimum support count is 2 and the final item set C3 will form after screening of the alarm item combinations in L2 (see Table 7).

The nonvoid proper subsets of $\{A1, A2, A3\}$ include $\{A1, A2\}$, $\{A1, A3\}$, $\{A2, A3\}$, $\{A1\}$, $\{A2\}$ and $\{A3\}$, and it can be inferred that the confidence coefficients are as presented in Table 8.

They meet the confidence coefficient confidence = 60% and the association rules are obtained $\{A1, A2\} \rightarrow A3$, $\{A1, A3\} \rightarrow A2$, $\{A2, A3\} \rightarrow A1$, $A1 \rightarrow \{A2, A3\}$; that is, Alarm3 will necessarily appear when Alarm1 and Alarm2 occur concurrently; the probability of concurrent occurrence of Alarm2 and Alarm3 is 67% when Alarm1 occurs; the rules for others are similar.

Based on the thinking of the Apriori algorithm flow above, the characteristics are as follows.

(1) Advantages: the algorithmic logic is clear without any complex mathematical derivation process with the dual parameter values of the support and confidence coefficient as the interest indicator for weighing the association rules.

(2) Disadvantages: frequent repeated scans of the information in the sample database lead to a heavy load on the system I/O; the number of the candidate frequent item sets increases sharply and the operation time increases significantly when the item sets are large; the attribute difference and importance of the set elements is ignored and high-value information is lost when the support and confidence coefficient serve as the sole criterion for weighing the item sets; the single-dimensional Boolean type association rules mining mode is used and multidimensional, multilevel, and numeric type association rules need to be improved.

In response to disadvantages of the Apriori algorithm, researchers compress the database samples by random sampling, formulate hash functions to the size of the candidate item set, reduce the number of scanning of the database by the method of dynamic item set counting, quickly establish frequent item sets utilizing the relation of “local-overall,” optimize the event database to reduce the quantity of the item sets in combination with the nature of the Apriori algorithm, use parallel computation, and so forth [14–16].

Based on the Apriori algorithm thought, Han et al., a professor from Simon Fraser University, adopted a partition search method combining expanded prefix tree data structure and branch-like local growth, that is, FP-growth (frequent pattern-growth) algorithm in 2000 [17], which avoids the problem of repeating an ergodic database in the Apriori algorithm and substantially improves the mining efficiency of association rules.

5.2. Particle Swarm Intelligence Algorithm. The adaptivity and high-efficiency characteristics of group system consisting of the natural ecosystem and various kinds of organisms in response to complex problems (e.g., community cooperation, biological evolution, immune system, nerve conduction, etc.) provide new research directions and application schemes

TABLE 7: Final item set of alarm item sets C3.

Alarm item	Support count
$\{A1, A2, A3\}$	2

for complex scientific problems, for example, ant colony algorithm, bat algorithm, bee algorithm, firefly algorithm, cuckoo search algorithm, particle swarm optimization algorithm, and so forth [18]. In 1987, the zoologist Reynolds simulated the process of aggregating and flying of bird flock self-organization by establishing flight rules for individuals of the bird flock, that is, collision avoidance, velocity matching, and flock centering [19]. In 1995, Kennedy and Eberhart analysed the process of aggregating, scattering, and migrating of birds; that is, when a bird flock searches for specific food in a certain unknown area at random, all individuals of the bird flock do not know their locations but they know the distance between their locations and the food. The simplest and efficient strategy is to search for the peripheral region of the bird closest to the food [20]. The whole foraging process achieved information sharing and competitive collaboration among individuals of the low-intelligence bird flock. In addition, the process embodies the value of the group intelligence evolving from unordered to ordered in obtaining the optimum solution. Kennedy considered the individuals of the birds as single particles and proposed the particle swarm optimization (PSO). The whole process follows the principles of environmental stimulus evaluation, adjacent individuals comparison, and learning adjacent advanced individual [21].

The PSO algorithm first initializes the particle swarm; that is, random location and velocity are assigned to the particle swarm in the feasible solution space. Each particle is a feasible solution in the optimization problem. A fitness value is determined by an optimization function; then, each particle will move in the solution space and the particle velocity will determine its motion direction and distance. Usually, particles approximate the current optimal particle until the optimal solution by means of iteration and each particle will approximate two optimal solutions during iteration, that is, particle optimum solution (POS) and global optimum solution (GOS).

5.2.1. Fundamental Principles of PSO. Assume a d -dimensional target search space, there is a group of particle swarms consisting of m particles with potential problem solution S , $s = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_m\}$, among which $\vec{X}_i = (x_{i1}, x_{i2}, \dots, x_{id})$, $i = 1, 2, \dots, m$ indicates a vector point of i th in the d -dimensional solving space; \vec{x}_i is substituted into the objective function pertinent to solving problem and the matched fitness value can be obtained. $\vec{P}_i = (p_{i1}, p_{i2}, \dots, p_{id})$ is used and $i = 1, 2, \dots, m$ indicates the optimum value point of the i th particle obtained by self-search (the optimum value means that its corresponding fitness value is the minimum); in the particle swarm S , there is an overall optimum particle, which is calculated as $\vec{G}_i = (g_{i1}, g_{i2}, \dots, g_{id})$, $i = 1, 2, \dots, m$; each particle also has a velocity variable

TABLE 8: Calculation of confidence coefficient of C3 alarm item set.

Logical relationship among alarms	Analytical calculation of confidence coefficient
$\{A1, A2\} \rightarrow A3$	Confidence = $2/2 = 100\%$
$\{A1, A3\} \rightarrow A2$	Confidence = $2/3 \approx 67\%$
$\{A2, A3\} \rightarrow A1$	Confidence = $2/3 \approx 67\%$
$A1 \rightarrow \{A2, A3\}$	Confidence = $2/3 \approx 67\%$
$A2 \rightarrow \{A1, A3\}$	Confidence = $2/4 = 50\%$
$A3 \rightarrow \{A1, A2\}$	Confidence = $2/4 = 50\%$

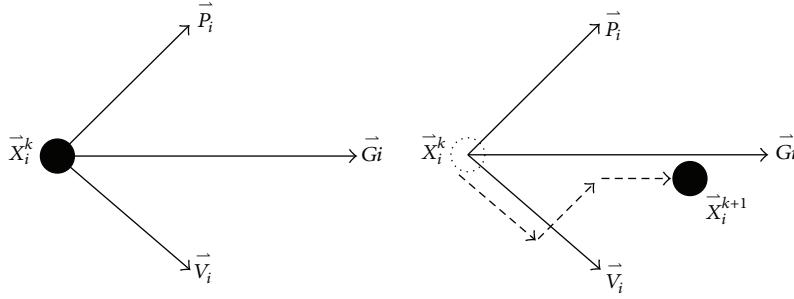


FIGURE 3: Particle migration of PSO.

$\vec{V}_i = (v_{i1}, v_{i2}, \dots, v_{id})$, $i = 1, 2, \dots, m$ indicating the velocity of the i th particle.

In the PSO algorithm, the following formulae are used for recursive calculation of particle movement:

$$\vec{V}_i^{k+1} = \vec{V}_i^k + c_1 * r_1 * \left(\vec{P}_i^k - \vec{X}_i^k \right) + c_2 * r_2 * \left(\vec{G}_i^k - \vec{X}_i^k \right), \quad (1a)$$

$$\vec{X}_i^{k+1} = \vec{X}_i^k + \vec{V}_i^{k+1}, \quad (1b)$$

where the particle number is $i = 1, 2, \dots, m$; k is the number of iterations; learning factors c_1 and c_2 are positive constants to which 2 is usually assigned; r_1 and r_2 are random numbers distributed between $[0, 1]$. In order to maintain the values of \vec{V}_i^k and \vec{X}_i^k within a reasonable regional range \vec{V}_{\max} and \vec{X}_{\max} should be set rationally.

Formula (1a) encompasses three facets of information when calculating the new velocity \vec{V}_i^{k+1} of the particle i : firstly, velocity \vec{V}_i^k is the speed of the particle i at the previous moment, secondly, information on distance between the current position of the particle i and the optimum position of the individual particle, and thirdly, the information on the current position of the particle i and the optimum position of the overall particle swarm. Formula (1a) is deployed to calculate the new position coordinates of particles. Formula (1a) and formula (1b) jointly determine the next motion position of the particle i . Taking a two-dimensional space as an example, Figure 3 describes the process where a particle moves from its initial position to its new position based on formula (1a) and formula (1b).

From the social dynamics, an analysis is conducted: the first part of formula (1a) is the memory term reflecting the velocity vector of particle in the previous step; the

second part is self-recognition term, a vector pointing to the optimum point of the particle from the current point, reflecting self-learning judgment of the particle under the effect of ambient particle swarm; the third part is the group-recognition term, a vector pointing to the optimum point of the overall particle swarm from the current point, reflecting experience sharing and collaboration among particles. The process reflects the basic learning development rules for biotic communities in the nature, that is, the process where companion knowledge learning and self-cognitive decision-making are integrating under constant action of external environmental information.

5.3. Optimization Algorithm for Mining of Particle Swarm Association Rules. Based on an analysis of the flow for the Apriori algorithm and particle swarm optimization, it has been discovered that the process of searching for the frequent items in the Apriori algorithm is actually a global search process while the particle swarm optimization is an algorithm finding the optimal global solution with excellent optimal performance. Therefore, the global optimum characteristic of the Apriori algorithm and the high efficiency of seeking the global optimal solution of the particle swarm optimization are needed for combining to achieve the optimization algorithm for association rules mining-APPSO algorithm.

5.3.1. Basic Flow of the APPSO Algorithm. The Apriori algorithm includes two stages and its overall performance is primarily determined by the first link, which aims at finding all frequent item sets meeting the minimum support in the database; the second link refers to finding the association rules meeting the minimum confidence coefficient from the frequent item sets.

Create three particle swarms in APPSO algorithm (see Figure 4), that is, the sample particle swarm, the candidate

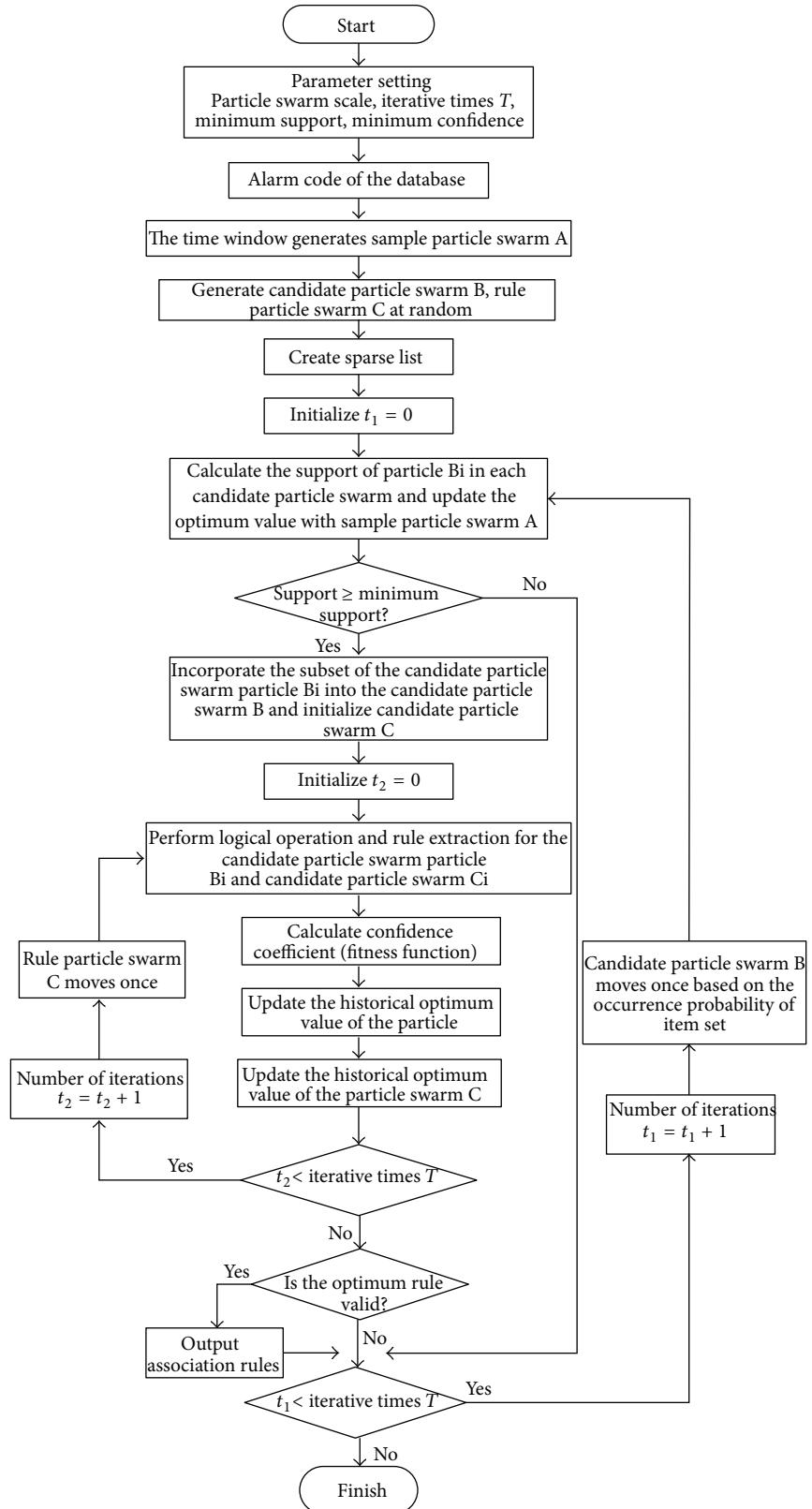


FIGURE 4: Basic flow of the APPSO algorithm.

particle swarm, and the rule particle swarm. The sample particle swarms are entity particle swarms; taking four-dimensional alarm data as an example, the sample particles are (A1, A3, A2), (A1, A2, A4); the candidate particle swarm and the rule particle swarm are logical particle swarms, for example, (1110) and (1101). The eligibility of the particles in the candidate particle swarm for candidate particles is determined by calculating and determining whether the particles in the sample particle swarm satisfy the minimum support. The particles in the candidate particle swarm and the rule particle swarm are judged logically to generate preliminary association rule. The association rules will be output if each preliminary association rule satisfies the minimum confidence; otherwise they will be discarded. The creating process is as follows.

(i) Sample particle swarm: the alarm data source is partitioned to create sample particle swarm A (SPS-A for short) by sliding the time window. For instance, after the number N time window capturing the natural time, the alarm sequence is shown in A1, A3, and A4; namely, the particle is A1, A3, and A4.

(ii) Candidate particle swarm: B particle swarm is created randomly in the APPSO algorithm (corresponding to the first link in the Apriori algorithm) such that each particle of the candidate particle swarm represents a certain candidate item set and all candidate particles of the whole candidate particle swarm represent a collection of all existing different candidate item sets. The support of the item set represented by each candidate particle is calculated to judge whether it meets the minimum support count value (calculation method, see Section 5.1.2). Such a particle swarm is referred to as candidate particle swarm B (Particle swarm CPS-B).

It is assumed that there are 4 types of alarms in the alarm database and they are Alarm A1, A2, A3, and A4, respectively. Each alarm is expressed with 0 or 1. 0 indicates that the alarm is not in the candidate particle currently while 1 indicates that the alarm is in the candidate particle currently. It is assumed that the value of a candidate particle is 1100; that is, Alarm A3 and Alarm A4 are not in the candidate particle and the particle represents a 2-item set consisting of A1 and A2. If the 2-item set meets the minimum support count value for sample particle swarm, the certain candidate particle would be reserved or removed conversely.

(iii) Rule particle swarm: in the APPSO algorithm, a particle swarm is randomly created (corresponding to the second link in the Apriori algorithm) such that each particle of the particle swarm represents a potential association rule. The length of each particle is equal to the length of each particle in the candidate particle swarm. Each alarm is expressed with 0 or 1. 1 indicates the corresponding alarm is the antecedent of the association rule while 0 indicates that the corresponding alarm is the consequent of the association rule. Such a particle swarm is referred to as rule particle swarm C (RPS-C).

Assume the value of a certain particle b in particle swarm C is 111000 and then the rule represented is $(A1, A2, A3) \Rightarrow (A4, A5, A6)$.

After creating of candidate particle swarm B and rule particle swarm C, the operational method for the two particle

swarms is as follows (particle a belongs to candidate particle swarm B and particle b belongs to rule particle swarm C).

The logic operation of “and” is performed for each particle of candidate particle swarm B and each particle of rule particle swarm C and the operational result is used to estimate the relation between the antecedent and consequent of the rule. For example, $a = 110011$, $b = 111000$, and $a \cap b = 111000$ indicate that Alarm A3 and Alarm A4 are not in the association rules. The field value of A2 and A2 is 1 and the field value of A4 and A6 is 0. We can obtain that the association rule represented by a and b is $(A1, A2) \Rightarrow (A5, A6)$.

5.3.2. APPSO Algorithm Optimization Link. During mining of association rules based on swarm intelligence, the particle ergodic method is usually used to obtain the support of the item set represented by the particle. The particle support obtained by scanning the whole database is accurate in result. However, some shortcomings exist; that is, the actual analysis efficiency is low and no data source characteristics and basic algorithm characteristics are combined. Therefore, data source sequencing coding and sliding window value assignment are used based on the data characteristics of the network alarms; the sparse linked list algorithm is deployed to calculate the support of the item set.

(1) *Sequencing Code.* As alarm names are usually described with English character string or digit combined number, such an identification method would bring about a large amount of combined data (e.g. MPLS_TUNNEL_MISMERGE and 007-061-00-800446) resolution consumption to data processing and analysing. Therefore, we employ the method by sequencing codes to reduce resolution consumption, in which all alarm names or network management alarm IDs are sequenced on the basis of the sequence of letters and figures. It targets on avoiding two or more integral values being assigned to the same alarm subsequently (Figure 5); differentiated values are assigned on the basis of data sequence.

(2) *Sliding Window.* Due to the combination of time-type data and relationship type in alarms, the time-type alarm data is sequenced on the basis of time length, the size of sliding time window, and sliding step length, and the relationship type alarm data is converted and combined into different transactional data item sets.

(3) *Sparse Linked List.* Compared with the overall alarm database after division, each of the alarm data item sets only contains partial alarm data types. The efficiency of database scanning by the APPSO algorithm is further improved using the thought of sparse linked list based on the data characteristics. The algorithm process is as follows.

A linked list header is created for each item of the whole database. For example, if there are 200 alarm code integer data types, consisting in 10,000 item sets, 200 linked list headers will be created and the integral value of each item is the number of its corresponding linked list.

The item sets are scanned in sequence and the items of each item set are added to the end of the corresponding linked list. For example, If the n th item set in the database is (50, 108,

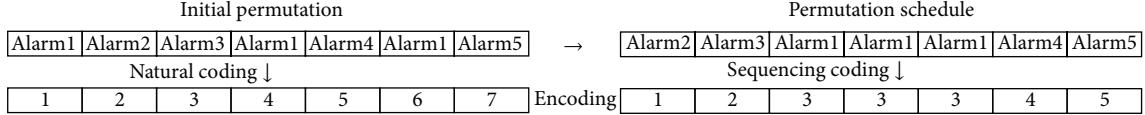


FIGURE 5: Natural coding and sequencing coding.

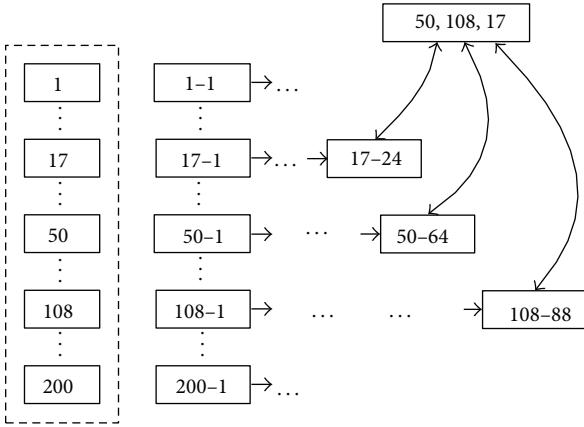


FIGURE 6: Examples of sparse linked list.

17), then the n th item set is added to the end of the linked list 50 and the end of the linked list 108, and so forth. 200 linked lists are created finally, that is, sparse linked list. The number of the alarm code integers saved in each linked list is much less than the 10000 item sets of the whole database (Figure 6).

(4) *Calculation of the Particle Support Based on the Sparse Linked List.* Take the n th item set in the database (50, 108, 17) and 200 linked list headers as examples (Figure 7).

Starting with the linked list 50, it is assumed to be containing the item “50” through searching the 64th item set. Similarly, the linked lists 108 and 17 correspond to 88 and 24, respectively; that is, all item sets before the 88th item set do not contain the corresponding item of the particle. After searching in the 88th item set, 1 will be added to the particle support if it contains (50, 108, 17) (Step 1), otherwise, continually searching in the linked list (50, 108, 17) in order to find the next data, respectively. Assume that they correspond to 121, 90, and 65, respectively, and directly search in the 121st item set. 1 will be added to the particle support if it contains (50, 108, 17) (Step 2); otherwise, continue to search in the linked list (50, 108, 17) and find the next data. Suppose that they correspond to 121, 184, and 121, respectively, and directly search in the 184th item set. 1 will be added to the particle support if it contains (50, 108, 17) (Step 3); otherwise, keep on searching. The overall linked list would finish searching when 50 has been sorted out in (50, 108, 17) (Step 4).

(5) *Nature of the Apriori Algorithm.* Based on the nature of the Apriori algorithm: “the subset of the known frequent item set k is also frequent;” the nature is used to optimize the

search rule for the particle swarm; that is, all subsets of the particle are also frequent if the corresponding candidate item set of a certain particle is a frequent item set. For example, if the particle a (110011) belongs to a frequent item set, then any subset of the value of a , such as 110000, 000011, 100001, 010010, 100010, and 010001, are frequent and these subsets are directly incorporated into candidate particle swarm A as new particles.

In conclusion, the main principle of the APPSO algorithm is to estimate whether each particle in candidate particle swarm A (CPS-A) is frequent or not. The subset of the particle will be added to A if the particle is frequent. Then the logical operation of “and” is performed for the particle and each particle of rule particle swarm B (RPS-B) to judge whether the corresponding rule of the result obtained is an association rule meeting the conditions or not. In accordance with a certain sequence, A and B are constantly updated until all iterative processes terminate.

5.3.3. APPSO Algorithm Test. A comparison test is conducted on the test platform with the APPSO algorithm and Apriori algorithm (hardware: CPU Intel, Core i5 3.3 GHz, 8 G RAM, 1 T hard disk, software: operating system window7, development platform Qt4.7.0, single-thread development). The alarm data (21084 pieces) of the network management system PTN device is extracted at random as the data. The data is generated into item sets with 5-seconds (5 s) time window and the data set containing only a single item (1-item sets) is rejected. Finally, 4753 item sets in total are obtained. The scales of candidate particle swarm and the rule particle swarm are identical.

(i) Test 1: relation between the support and number of association rules: the scale of the particle swarm is 40, number of iterations is 100, and confidence coefficient is 30%.

Analysis on Test 1: Apriori algorithm is a global search algorithm. Therefore, the number of the association rules mined by the APPSO algorithm is less than the number of the association rules mined by the Apriori algorithm. More than 60% of the main association rules is obtained with the APPSO algorithm as shown in Figure 8.

(ii) Test 2: relation between the confidence coefficient and number of association rules: the scale of the particle swarm is 40, number of iterations is 100, and confidence coefficient is 5%.

Analysis on Test 2: under the condition of a constant number of iterations and minimum support, the number of alarms obtained by the two algorithms will necessarily decrease with increasing of confidence coefficient index; compared with the Apriori algorithm, when the confidence

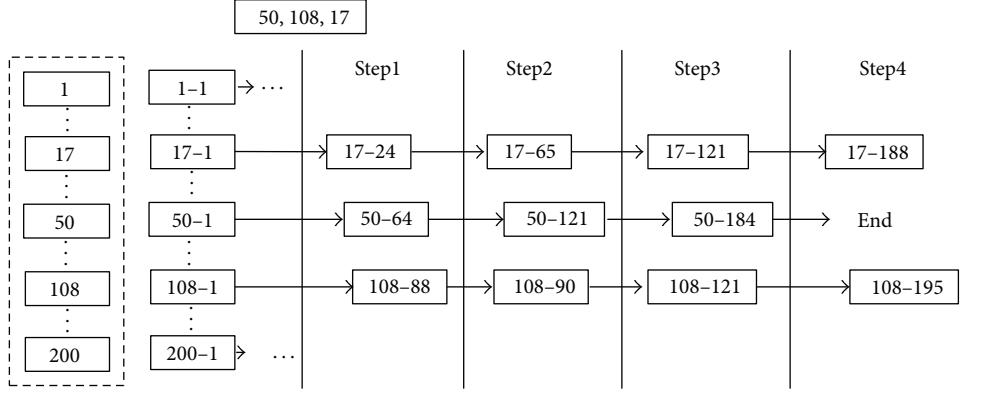


FIGURE 7: Examples of calculation of the particle support based on the sparse linked list.

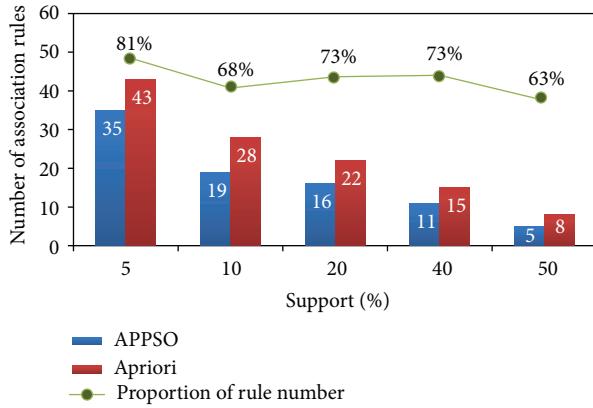


FIGURE 8: Relation between the support and number of association rules.

coefficient value is within the discrete interval [30%, 60%], the number of the association rules obtained with the APPSO algorithm accounts for approximately 80% as shown in Figure 9.

(iii) Test 3: relation between the scale of the particle swarm and the number of association rules: the number of iterations is 100, the minimum support is 5%, and confidence coefficient is 30%.

Analysis on Test 3: under the condition of a constant number of iterations, minimum support, and confidence coefficient, the larger the particle swarm is, the more the number of the association rules will be. The number of the association rules will approach the number of the rules obtained by the global search of the Apriori algorithm as shown in Figure 10.

(iv) Test 4: relation between the number of iterations and operation time: the scale of the particle swarm is 40, minimum support is 5%, and the confidence coefficient is 30%.

Analysis on Test 4: under the condition of a constant particle swarm scale, minimum support, and confidence coefficient, the time for the APPSO algorithm is prolonged

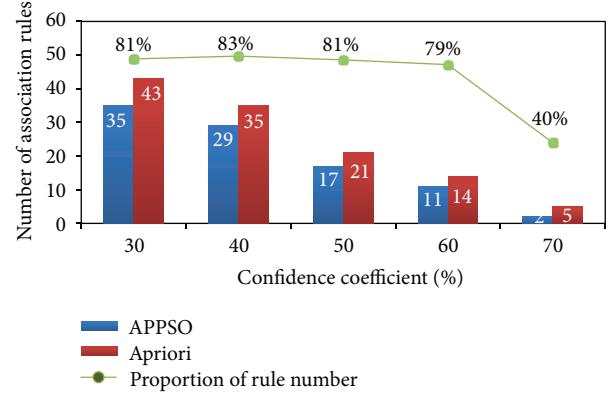


FIGURE 9: Relation between the confidence coefficient and number of association rules.

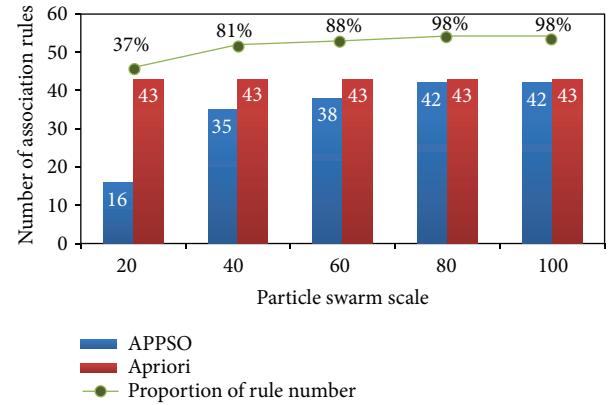


FIGURE 10: Relation between the scale of the particle swarm and the number of association rules.

with increase of the number of iterations but the number of association rules obtained significantly increases; compared with the Apriori algorithm, the efficiency of the APPSO algorithm significantly increases; for example, the number of iterations is 120, the time for the APPSO algorithm only

TABLE 9: Distribution of training data sets.

Item sets	1-item sets	2-item sets	3-item sets	4-item sets	5-item sets	6-item sets	7-item sets	8-item sets	9-item sets	10-item sets
Number of item sets	136342	91787	36780	10141	2578	626	206	109	188	204
Item sets	11-item sets	12-item sets	13-item sets	14-item sets	15-item sets	16-item sets	17-item sets	18-item sets	20-item sets	100-item sets
Number of item sets	194	704	343	239	186	54	30	6	0	1

TABLE 10: Distribution of test data sets.

Item sets	1-item sets	2-item sets	3-item sets	4-item sets	5-item sets	6-item sets	7-item sets	8-item sets	9-item sets
Number of item sets	15455	8649	1386	232	33	5	5	4	8
Item sets	10-item sets	11-item sets	12-item sets	13-item sets	14-item sets	15-item sets	16-item sets	17-item sets	18-item sets
Number of item sets	1	7	55	10	10	9	4	1	1

TABLE 11: Statistics on association rate of test data.

Minimum support count	0.01	0.01	0.01
Minimum confidence	0.01	0.05	0.09
Number of rules from training data sets	185	149	154
Alarm association rate from test data sets	81.23%	81.21%	81.22%

accounts for 17% of the time for the Apriori algorithm yet the number of the rules obtained accounts for 88% of the total number of rules as shown in Figure 11.

On the premise of desired demand for the number of rules, the APPSO algorithm is able to control the operational

precision and decrease the computation time and memory consumption by reasonably setting the particle swarm parameters.

(v) Engineering test: the network alarm data over the four of the 8 consecutive weeks is used as "training data." The alarm association rules are mined by the APPSO algorithm and the data over the other 4 weeks is used as "test data" to calculate the alarm correlation rate. Specific method: all alarms are intercepted as per the fixed flow time window and all of the non-1-item sets are included in the calculation of the alarm correlation rate (the 1-item sets themselves do not have a correlation relationship). The algorithm is as follows:

$$\text{Alarm association rate} = \left(\frac{\text{number of non-1-item sets meeting the association rules}}{\text{number of all non-1-item sets}} \right) \times 100\%. \quad (1c)$$

For example, the alarm sequence is (A1, A2, A3, A1, A4, A2, A3, A4, A4, A2), and becomes {A1, A2}, {A3}, {A1, A4, A2}, {A3, A4}, {A4, A2} after being intercepted in accordance with the fixed flow time window, among which the non-1-item sets involving in the calculation of the alarm correlation rate are {A1, A2}, {A1, A4, A2}, {A3, A4}, and {A4, A2}. The association rate of the alarm data is 50% if the association rule is A1 → A2.

Analysis on engineering test: The alarm association rules obtained through the training data over the first 4 weeks is applied in the test data over the last 4 weeks. The training data over the first 4 weeks contains the equipment types, BSC, BTS, CELL, and 516271 alarms, of which the alarm types are 131. The time window is set to 2 s and the sliding step length to 1 s; the test data over the last 4 weeks contains the equipment types, BSC, BTS, CELL, and 39470 alarms, of which the alarm types are 89. In combination with the requirements for actual

conditions of the engineering operating environment, the time window is set to 3 s. 10420 non-1-item sets are obtained after interception of data.

From Tables 9, 10, and 11 it is obtained that all of the alarm association rates are higher than 80%. The APPSO association mining algorithm provides an effective analytic method for the alarm association analysis.

6. Conclusion

The association rules for the alarm data in the information communication network should be analysed in conjunction with the data characteristics to perform a design specifically to achieve a corresponding algorithm flow. Compared with the Apriori algorithm, the mining efficiency of the APPSO algorithm is significantly enhanced but a small number of

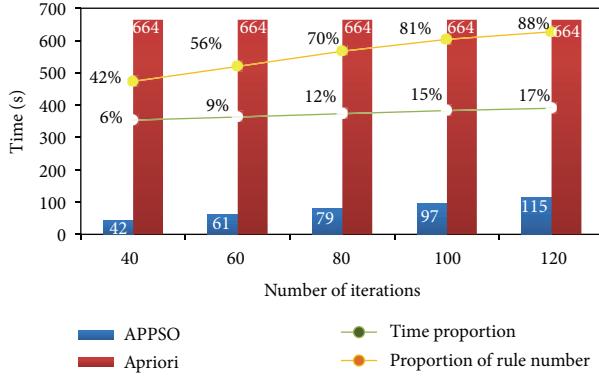


FIGURE 11: Relation between the number of iterations and operation time.

association rules are lost to some extent due to the characteristics of the PSO algorithm. The value of the association rules lies in quick acquisition and subsequent high-value evaluation of association logic instead of sole acquisition of all association rules. From this perspective, the APPSO algorithm improves in both mining efficiency and algorithm concept.

Conflict of Interests

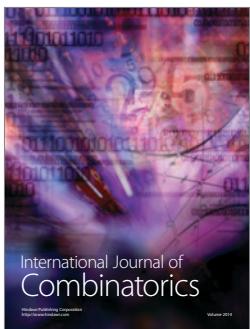
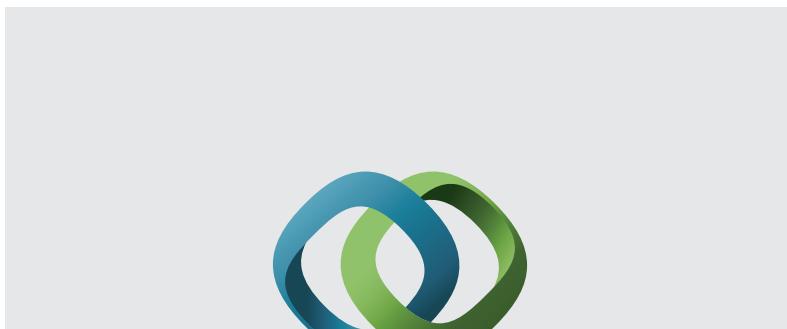
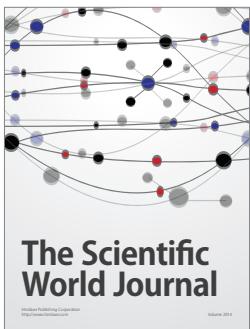
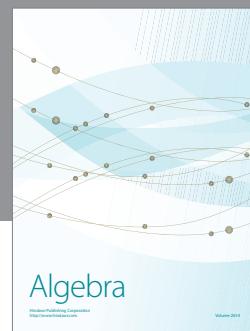
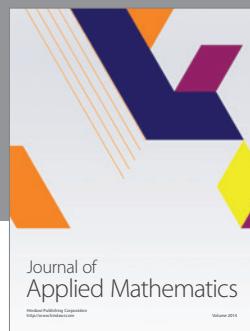
The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research was supported by a grant from the National Natural Science Foundation of China (no. 51205274), the Science and Technology Major Project of the Shanxi Science and Technology Department (20121101004), the Key Disciplines Construction in Colleges and Universities of ShanXi ([2012]45), the Shanxi Scholarship Council of China (no. 2013-035), the China Postdoctoral Science Foundation (no. 2013M530894), and the Innovation Project of the Postgraduate Education in Shanxi Province (no. 20123027).

References

- [1] E. Kiciman and A. Fox, "Detecting and localizing anomalous behavior to discover failures in component-based internet services," Tech. Rep., The Stanford Computer Science (CS) Department, Stanford, Calif, USA, 2004.
- [2] L. L. Huang, G. G. Su, and Y. J. Jiang, *Operation Support System Technology and Practice*, Posts & Telecom Press, Beijing, China, 2012 (Chinese).
- [3] D. T. Li, *Researches on data mining based alarm correlation analysis in communication networks [Ph.D. thesis]*, University of Electronic Science and Technology of China, Chengdu, China, 2010 (Chinese).
- [4] B. Z. Yao, R. Mu, and B. Yu, "Swarm intelligence in engineering," *Mathematical Problems in Engineering*, vol. 2013, Article ID 835251, 3 pages, 2013.
- [5] Y. Wang, G. C. Li, and Y. K. Xu, "Research on management method, classification and correlation of alarm in information communication network," *Telecommunications Science*, vol. 29, no. 8, pp. 132–135, 2013 (Chinese).
- [6] X. B. Wang, W. Li, and H. W. Xu, "Management analysis of alarm standardization in centralized operational mode," *Telecommunications Technology*, no. 4, pp. 39–42, 2009 (Chinese).
- [7] T.-Y. Li and X.-M. Li, "Preprocessing expert system for mining association rules in telecommunication networks," *Expert Systems with Applications*, vol. 38, no. 3, pp. 1709–1715, 2011.
- [8] H. Mannila, H. Toivonen, and I. Verkamo, "Discovery of frequent episodes in event sequences," *Data Mining and Knowledge Discovery*, vol. 1, no. 3, pp. 259–289, 1997.
- [9] R. Sterritt, D. Bustard, and A. McCrea, "Autonomic computing correlation for fault management system evolution," in *Proceedings of the IEEE International Conference Industrial Informatics (INDIN '03)*, pp. 240–247, Alberta, Canada.
- [10] A. A. Amaral, B. Z. Zarpelão, L. M. Mendes et al., "Inference of network anomaly propagation using spatio-temporal correlation," *Journal of Network and Computer Applications*, vol. 35, no. 6, pp. 1781–1792, 2012.
- [11] X. D. Wu and K. Vipin, *The Top Ten Algorithms in Data Mining*, Chapman and Hall/CRC, Boca Raton, Fla, USA, 2009.
- [12] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases," in *Proceedings of the 20th International Conference on Very Large Data Bases*, pp. 487–499, Santiago de Chile, Chile, 1994.
- [13] S. Y. Jiang, X. Li, and Q. Zheng, *Principles and Practice of Data Mining*, Publishing House of Electronics Industry, Beijing, China, 2011 (Chinese).
- [14] T. Calders, N. Dexters, J. J. M. Gillis, and B. Goethals, "Mining frequent itemsets in a stream," *Information Systems*, vol. 39, pp. 233–255, 2012.
- [15] V. D. Mabonzo, *Study on new approach for effective mining association rules from huge databases [Ph.D. thesis]*, Dalian Maritime University, Dalian, China, 2012.
- [16] K. Z. Ziauddin, K. T. Shahid, and Z. K. Khaiuz, "Research on association rule mining," *Advances in Computational Mathematics and Its Applications*, vol. 2, no. 1, pp. 226–236, 2012.
- [17] J. W. Han, J. Pei, and Y. W. Yin, "Mining frequent patterns without candidate generation," in *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD '00)*, pp. 1–12, Dallas, Tex, USA, 2000.
- [18] X. S. Yang, Z. H. Cui, R. B. Xiao et al., *Swarm Intelligence and Bio-Inspired Computation: Theory and Applications*, Elsevier, Amsterdam, The Netherlands, 2013.
- [19] C. W. Reynolds, "Flocks, herds, and schools: a distributed behavioral model," *Computer Graphics*, vol. 21, no. 4, pp. 25–34, 1987.
- [20] J. Kennedy and R. C. Eberhart, "Particle swarm optimization," in *Proceedings of the IEEE International Conference on Neural Networks*, pp. 1942–1948, December 1995.
- [21] G. Veysel and M. P. Kevin, *Swarm Stability and Optimization*, Springer, Berlin, Germany, 2011.



Submit your manuscripts at
<http://www.hindawi.com>

