

Alarm Correlation Analysis in SDH Network Failure

Zhen-dong Zhao

Department of electric and
communication engineering
North china electric power university
Yonghua Street 619,071000 Baoding,
China
bdzzd@sina.com

Huang Nan

Department of electric and
communication engineering
North china electric power university
Yonghua Street 619,071000 Baoding,
China
huangnan1987@126.com

Zi-han Li

Department of electric and
communication engineering,
North china electric power university
Yonghua Street 619,071000 Baoding,
China
purplewinter@126.com

Abstract—According to the characteristics of alarm data and the alarm spread in SDH system, and the existence of various shortcomings of the classical algorithm, this paper presents the FP-Growth of non-temporal association rule mining algorithm. We join the relationship between network topology as a constraint to filter the interference of unrelated network equipment alarms form the related network element alarms. In the process of alarm association rule mining we improve the timeliness and accuracy of the results.

Keywords—Alarm correlation; Network topology constraints; FP-Growth algorithm; Data mining

I. INTRODUCTION

As an advanced science and technology in today's communications field, SDH optical transmission has been widely used in the power system. How to do the daily operation and maintenance work of SDH equipment in power communication effectively and ensure their safe and stable operation are the important tasks for operation and maintenance personnel. When the network fails, it often requires that we must judge the location, type and reason of network failure in the shortest possible time so as to repair the fault in a timely manner [1].

A lot of historical alarm data is stored in the database of network alarm, this alarm data either is incomplete or contains more redundant information, but it contains some valuable information [2]. We can use association rule mining techniques to dig out the potential and useful rules. When real-time alarm data matches with these rules, we can deal with the fault with the corresponding program.

Currently, most of mining for SDH system alarm correlation ignores the global information of network topology, which makes the data to be mined very large and mines too many association rules with low accuracy [3]. To solve the problem, this paper proposes an association rule mining algorithm using the constraints of the network topology so as to improve the timeliness and accuracy of the results in the process of alarm association rule mining.

II. PRETREATMENT

The original data is collected from the network management system, so it inevitably contains with some noise and redundancy. The pretreatment of data to be mined is an

important prerequisite and step for data mining. According to alarm features, we can clean up the noise of raw data with different filtering methods. For example, we can filter the alarms (flash off) with the same start and end time, and filter the alarms of specified network element or the specified time period, and we also can establish the filtered alarm list. We remove redundancy in the raw data through the alarm compression, and retain only the useful data for the follow-up data mining algorithm. We compress primarily the repeated alarms produced before the fault is cleared, and take only one of the repeated alarm records.

The input format of the data mining algorithm is transactional data, and the original alarm data is relational data, so we need to do the data conversion. In the conversion process we must divide the records of original alarm information with a reasonable rule, and organize reasonably these records to generate a transaction set. Generally the traditional sliding window [4] partition method has overlap, and the overlapping range depends on the sliding step size accounted for the proportion of the size of the window. In addition, it maybe divides the related alarms into different sequences. It will generate a lot of empty window (without the alarm window) when the window slides.

Because of the complexity of the network itself and its physical and logical correlation, single fault may lead to a large number of alarms, and form the so-called "alarm storm". These alarms triggered by the same failure occur inevitably at the same or similar time. That is to say, the alarms with the similar start time may have relationship, and those with longer time interval may not be caused by the same fault. In the original alarm data, we have detected "long enough" time periods without alarms. Those periods are considered as cluster borders. All alarms suited between two cluster borders belong to the same cluster [5]. This method may lead to a relatively large number of affairs, in other words, the amount of data to be mined will be very large. And we will get substantial and redundant association rules. Therefore, we have to build the network topology model, and filter the network element alarm without relationship in the same transaction to improve mining efficiency and correctness.

III. IMPLEMENTATION

A. Network topology constraints

SDH transmission network is made up of various network elements, the basic types of network elements is TM, ADM and SDXC etc. Transmission network refers to the physical network formed by the actual information transmission equipment (such as fiber optic cable). Analyze the network structure of the SDH system, use a database table to create and store network element topology, and achieve the modeling of the physical network topology. The topology model is an abstract view of network structure, which hides other aspects having nothing to do with the network topology. We can understand the connection among the various network elements intuitively, and fine the characteristics of the network connection.

In preprocessing phase, for an alarm of the original alarm data to be analyzed, find out all or most of the network elements having topological relations with it. Extract these network elements connected together, and store them in some kind of data structure in the form of network elements, forming a network element cluster. Then match the original alarm data in the same transaction set with the network element clusters, and filter the alarms not belong to the clusters. This reduces the width of the transaction, and reduces the workload significantly. At the same time we ensure that the rules meet the topological relations.

B. FP-Growth algorithm

Many researches of association rule mining algorithm are apriori algorithms or derived algorithms based on apriori algorithm. Apriori algorithm uses the interaction of candidate item sets and frequent item sets, gets all frequent item sets, and greatly reduces the size of candidate item sets through the pruning of candidate item sets. When the mining object has a large number of frequent patterns, long patterns, or the minimum support threshold is small, Apriori algorithm will face difficulties because of huge memory consumption and time-consuming mining process. Apriori like algorithms [6] have an inherent weakness, which generates frequent item sets relying on the candidate item sets, so they have no substantive breakthrough in the application. Jiawei Han proposed a new method using a compressed data structure (FP-tree) to store all the data needed in association rules mining. It scans the source data twice, and stores data in the FP-tree structure. It greatly reduces the overhead of data exchange and frequent match, avoiding to generate candidate item sets. It is the so-called no candidate set generation algorithm (FP-Growth). Yang Wei and others have compared Apriori and FP-Growth [7]. The result is as Figure 1.

1) Construct FP-tree

Scan the database once, get the frequent item sets F and its support, order F in descending according to the support, and produce frequent item sets L.

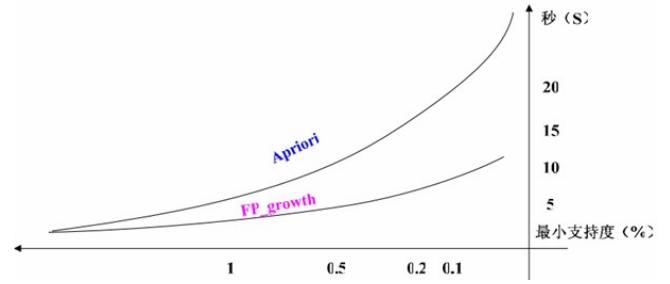


Figure 1 minimum support and the execution time ratio relationship

FP-Growth algorithm is described as follows [8]:

Create the FP-tree's root node, labeled null. For each transaction in the database, select the frequent items, and order them according to L. Suppose frequent entry table after sorting is $[p/P]$, where P is the first element, and T is the remaining elements of the table. Call the insert-tree $([p/P], T)$. The implementation of the process is as follows: If T has a child N, $N.item-name = p.item-name$, then N count plus 1; otherwise, create a new node N, set its count to be 1, link to its parent node T, and link to the node with the same key name through the node chain structure; if P is not empty, call insert-tree (P, N) recursively.

2) Mining Frequent Patterns in the FP-Tree

Input: the FP-Tree of transaction database D and the minimum support threshold minsup.

Output: a collection of all frequent patterns. Methods: call the FP-Growth (FP-tree, null).

```

Procedure FP-Growth (Tree,  $\alpha$ )
if (Tree contains only a single path P) then
    for each combination of nodes in the
    path P (denoted as  $\beta$ )
        generate pattern  $\beta \cup \alpha$  ,
        Support=Minimum support for all nodes in
         $\beta$ 
else on the Tree head each ai do
    generate pattern = ai, support =
    ai.support;
    construct conditional pattern database
    of  $\beta$  and its conditional FP-Tree Tree  $\beta$ ;
    if Tree  $\beta \neq \phi$ 
        then call FP-Growth (Tree  $\beta$ ,  $\beta$ )
    
```

This is arecursion calling function. According to the properties of the FP-tree and the above lemma and corollary, the algorithm can obtain all frequent item sets meeting the conditions for a given data source and support threshold.

The overall solution architecture diagram of fault management based on network topology constraints and FP-Growth algorithm is given as follows:

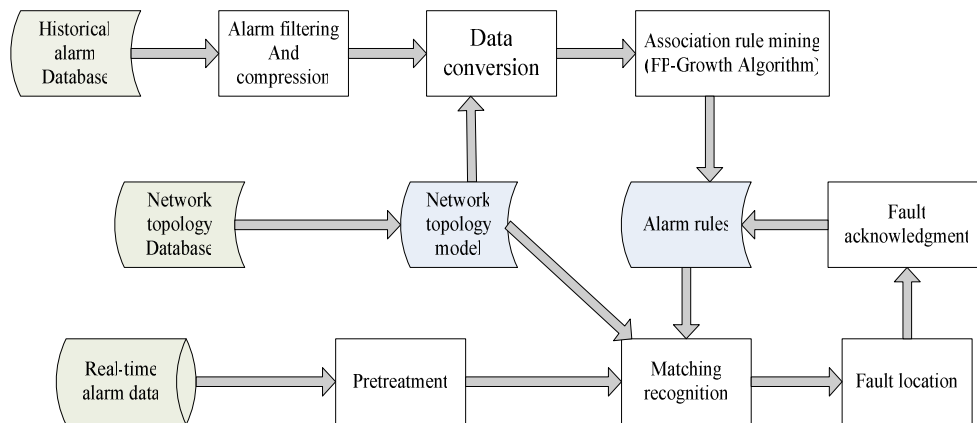


Fig 2 the overall solution architecture diagram of fault management

IV. CONCLUSION

According to the characteristics of alarm data and the alarm spread in SDH system, and the existence of various shortcomings of the classical algorithm, we presents the FP-Growth of non-temporal association rule mining algorithm. The algorithm only scans the database twice, and does not produce candidate sets, with a clear performance advantage. It can save a lot of time and improve efficiency in the implementation, especially when we face large databases. In the implementation process of the FP-Growth algorithm, we join relationship between network topology in the data conversion as a constraint. This step makes the final results conform to the actual correlation, avoid generates a lot of redundancy and meaningless rules, and improves the timeliness and accuracy of the results in the process of alarm association rule mining.

REFERENCE

[1] Wu Yang-yang, Chen Huai-nan. An Association-Rule-Based Model for

- Telecom Alarm Correlation Analysis[J]. Journal of Communication and Computer, 2004, 1(1):57~63.
- [2] Jiang Hong-wei, Yan Peng-bo. Customer Network Management Alarms Information Analysis Based on Transmission Network[J]. Communications Management and Technology, 2010, 8(4):38-41,47.
- [3] Wang Zhao-gang, Research and Application on a Topological Constraint Based Sequential Data Mining Approach [D]. 2009
- [4] Zhang Xian-fei etc. Analysis of power communication network alarms correlation based on incremental mining algorithm[J]. International Electronic Elements.2008, (10):67-68,71.
- [5] Oliver Jukic, Marijan Kunštic,ABCDE – Alarm Basic Correlations Discovery Environment [J]. MIPRO 2010, May 24-28, 2010, Opatija, Croatia.
- [6] ZOU Dong-sheng, Cheng Dai-jie. The Alorithm of Dataming Technology in Telecommunication Network Alarm[J]. Microelectronics and Computer.2001, (5):32-35.
- [7] Yang Wei etc. Comparision and Application of FP_growth and Apriori Algorithm. <http://www.paper.edu.cn>
- [8] Zheng Yan. Data warehouse and data mining[M].Beijing: Tsinghua University Press,2011.174-175.