Fault Mode Analyze of Power System Based on Data Mining

He YueShun1,2, Ding QiuLin1
1 Nanjing University of Aeronautics and Astronautics, Nanjing, China
2 East China Institute of Technology, Nanchang, China
e-mail:hys8418@163.com

Abstract—On power system operation status monitoring, operating performance analysis and assessment is to ensure the safe operation of its important components. This paper presents a new type of data mining based on the fault mode analysis and the fast diagnostic reasoning algorithm. Fault appearance to be collected and cleaned up in a fault information dimension table, the relationship rule dimension table was made up by the operating characteristic values and cause of the malfunction and other components. The data mining and analysis was implemented between the fault information dimensions table and the relationship rule dimensions table. We made sure the causes of the fault and chose the priority solution for troubleshooting by generating candidates sets and filtering the candidate set and matching the fault.

Index Terms—fault mode analyze; data mining; frequent item set; fault match; power system

I. POWER SYSTEM STATE MONITORING AND FAULT DIAGNOSIS STUDY IN THE PRESENT SITUATION

Power system goes wrong in randomness, when each fault has been taken places, it’s changes in the scope of its parameters have a very strong randomness. Power system state monitoring and fault diagnosis study of applied research at Present, there are more techniques and methods would be used, for example: expert systems, causal maps, fuzzy sets, inductive learning, Artificial Neural Networks, wavelet transformation, Kalman Filters, etc....In addition, the Chaos and Fractal Theories already begun to attract people’s attention. The traditional relay protection is based on the concept of steady-state, it required to refine fundamental component of fault signal in accurate, but the research is less based on fault transient signal component; When the power system go wrong, there are a large number of high-frequency transient state component in voltage and electric current, The size and distribution of transient state component are change with the fault time, fault location, fault point arc resistance, the system operating conditions change, transient signal caused by fault is a non-stationary random process, and the data mining is to analysis the fault, which based on high-frequency transient state component of fault, and which to predict, determine and deal with the possible fault, so that we can set up the scheduling plan for reasonable and guarantee the security of electricity supply and improve operation efficiency.

II. FAULT MODE

Power system in the long-term operation and maintenance has accumulated rich experience and a lot of original information, and it is important for enterprises to change these experiences and information into knowledge. Fault mode is the genuine cause of the fault which is the appearance summarized when the products take place fault. Fault mode is the one of the main target of data mining, the appearance of fault are classified and cleaned up to form the fault mode is a valuable resource for enterprises. The user's description of fault is often t he fault's appearance, and the fault appearance is only the outward manifestation of the issue of fault in many cases, the appearance of fault will correspond to the fault cause can be cleaned up and analyzed by the experience

<table>
<thead>
<tr>
<th>ID</th>
<th>Fault properties</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Fault code</td>
<td>The fault of unique identifier</td>
</tr>
<tr>
<td>2</td>
<td>Fault Statistics Code</td>
<td>Used for statistical analysis, the user can determine the statistics particle size</td>
</tr>
<tr>
<td>3</td>
<td>Fault Level</td>
<td>Fault Level can be divided into mild fault, general fault, a serious fault and a fatal fault</td>
</tr>
<tr>
<td>4</td>
<td>fault Parent code</td>
<td>The structure of fault mode organized by tree, and definite the logical relationship each other</td>
</tr>
<tr>
<td>5</td>
<td>Type of coding parts</td>
<td>Identification of the components correspond to the type of fault mode</td>
</tr>
<tr>
<td>6</td>
<td>Service time</td>
<td>The fault corresponding to the general maintenance services time</td>
</tr>
<tr>
<td>7</td>
<td>Cause of fault</td>
<td>The essence reasons for fault</td>
</tr>
<tr>
<td>8</td>
<td>Fault appearance</td>
<td>The reasons of the fault corresponds to the appearance of common faults(Fault manifestation)</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
and relevant technical standards and come into a standard dimension table of the fault mode. A typical fault mode recorded content as shown in table 1:

Among them, the items 1,2,4,7 are the key elements of the system and is the basic to deal with the fault information. The data of the fault mode dimension table initialization comes from the historical experience of enterprise data; Fault mode dimension table can be maintained by automatically and artificial maintenance, the main source of information is the user feedback of the new fault appearance.

III. DATA MINING -BASED FAULT DETECTION

A. The Gross Structure Design

In recent years, data warehouse technology is a better application, some enterprises set up their own data warehouse and have accumulated a large amount of data, however, data warehouse to be used as a database in some enterprise, data warehouse has not play an important role in effect. Therefore, people began to study in order to find useful knowledge from the huge data warehouse and to improve the utilization of information. A large number of valuable knowledge would be found from fault information for production, management and for maintenance equipment and circuitry to use for reference, so that, the accuracy of fault diagnosis is improved and the time of fault remove is shorten. Data mining process in the fault information was shown in Fig 1.

The main objective of data mining is to use the correlation of the fault information in data warehouse in order to find the information which is hidden and divinable and understandable and valuable, and to find the model which is easy understand by people to describe the data in the data warehouse. The data was collected according to fault model which was cleaned up and a fairly standard, it is easy for enterprises to establish knowledge dimension table and the relationship rules dimension table, in the data warehouse can be found a large number of knowledge about equipments performance, operation status, scheduling scheme, the cause of the fault, service decision, etc..., the knowledge play the role of reference and guidance for continuous improving circuit design, process management, power load forecasting, power market analysis, improving the quality of service in enterprise.

B. Data mining rules

The knowledge characteristics of diagnosis field and the equipment fault mode analysis are provided by a variety of technical parameters or experience, it combined with fuzzy logic methods, production rules based on the uncertain knowledge reasoning were adopted, the general form of the rules as follows:

IF P THEN Q, in which P = P1 (C1) AND / OR P2 (C2) ... Pn (Cn), P is fault appearance, it could be a simple fault appearance, also, it could be a logical combination by a number of a simple fault appearance, for example, P = P1 (C1) AND P2 (C2) OR P3 (C3); For each fault appearance P1, P2, ..., Pn can be assigned a corresponding confidence value C1, C2, ..., Cn, it is the percentage of the credibility of the fault appearance (probability of the fault appearance); Q is the real reason for the fault ,and it may be one or more conclusions. A new type of reasoning mechanism was adopted in the fault mode analysis which is inexact matching based on data mining, and it searches the rules which matches the terms in line with the known, It can generate many different fault conclusions, it is fit for multi-fault diagnosis of intelligent systems. When the quantitative analysis was given prominence to, some knowledge of qualitative judgments were strengthen or weaken the results of quantitative analysis by intelligent systems of highly interactive.

The support plays a decisive role in the Apriori algorithm; the conviction was the first place and gave up the definition of minimum support in the solution. The whole algorithm is divided into three steps: matching fault pattern, generating candidate item sets, filtering the candidate item sets. After “fault mode extraction” in the System, a code aggregate (dynamic string array Arr1) was generated which included the fault reasons.

C. Matching fault mode, to identifying the frequent itemsets about the cause of the fault

1) The code aggregate of the cause of fault was sorted and generated " Fault_Model":

Package1:First of all, the sequence of the fault dimension table was created in the data warehouse, the corresponding ID number was found according to these faults number, and the ID numbers were sorted, an array Arr2 was created, it’s size as same as the array Arr1.Then the output results of the Package1 were be converted into numeric, and it kept in the computer's memory as hexadecimal number.
2) A "Test Model" was created by the rules of the dimension table records, and with the "Fault Model" match, then an ideal frequent item sets and candidate frequent item sets were generated (Note: Frequent Item sets 1: ideal frequent item sets; frequent item sets 2: the candidate frequent item sets; frequent item sets 3: callback frequent item sets). The contents of the first record were read in Arr1 when the qualification was not empty in the rules dimension table, executing Package1, then the results were saved in another variable "Test_Model".

3) We would be carried out with “Fault_Model” and “Test_Model” computing as the rules of data mining were given in previous, the frequent item sets 1 and frequent item sets 2 were generated.

At last, traversing each records in the rules dimension table, repeat it, the frequent item sets 1 and frequent item sets 2 were generated. If the frequent item sets 1 is empty while it replaced by the contents of frequent item sets 2, while empty frequent item sets 2.

D. Filter candidate itemsets and decide the fault type

Matching fault mode combine with the fuzzy inference rules to filter the dimension.

1) When the frequent item sets 2 is empty, the more of the special rules and the fault number are priority in rules of the frequent item sets 1, the aggregate of conclusion’s number could be got immediate in this fault diagnosis by the frequent item sets 1. The output was a single conclusion if there was only one element in the frequent item sets 2, the fault was a single fault; Otherwise, the output was more conclusions and the fault was blend faults, more of the fault of electrical power equipment and circuits were blend faults.

2) When the frequent item sets 2 is not empty, the first, all the rules were deleted if the number of qualification less than or equal the rules of the maximum number in frequent item sets 1; If frequent itemsets2 is not empty, the ID of rules which was the maximum number of rules for the fault was added in frequent itemsets1( the ID of rules were preserved and the others were removed if there are special rules), at the same time, frequent itemsets2 was deleted, but the rules were signed and separated for the original ideals and candidates frequent item sets, among them, the exceptional fault ID were recorded for the rules of the candidates frequent item sets (failed to match the fault ID), it is a reference for users.

3) If the frequent itemsets1 and frequent itemsets2 are empty, then goto the display program for diagnosis directly, it showed that there no rules match, the system can do nothing for the fault diagnosis, and it can be identified and estimated only by manual.

IV. CONCLUDING REMARKS

This paper studies the application with the data mining in power system state monitoring and fault diagnosis, presents a new type of data mining based on the fault mode analysis and the fast diagnostic reasoning algorithm, it has obvious advantages in dealing with a large number of power system data. In combination with other data mining algorithms, we can further improve the system state analysis and mining; experiment proved that the method is particularly fit for dealing with the discrete or logic data of the power system. We can effectively decide the operating state of power system and the improve the accuracy of the fault mode analysis. After the system was implemented; The data was analyzed by a logical of the relevance degree, We can improve the quality of service of electric power enterprises and shortened the service time; at the same time, it could provide a great deal of information for aid decision making for planning and designing new electric power for enterprises.

REFERENCES


