

A Study on the Application of Defect Data Mining in the Aid Decision Making of Dispatching and Control Integration

Hongzuo Guo, Huayong Lu & Haitao Su
State Grid Henan Electric Power Company, Zhengzhou, Henan, China

Lifang Li & Sheng Li
Tellhow Soft Co., Ltd, Nanchang, Jiangxi, China

ABSTRACT: The purpose of the research project is to design an aid decision-making system of dispatching and control integration that cooperates with expert judgments based on the data mining of equipment defects so as to improve the risk pre-warning and pre-control ability of unattended centralized monitoring, enhance the level of centralized monitoring operation and guarantee the safety and stability of power system in the dispatching and control integration system. Defect analysis and condition based maintenance are deeply integrated in this research project. Condition data of equipment defects is firstly obtained and then classified, aggregated and analyzed. Through the model training of SVM, the law of the big data when equipment defects occur is obtained and the operation curve of equipment and the current status information are matched so that the risk of defects can be confirmed and the accuracy of risk anticipation can be effectively improved. In this research project, familial defects of equipment and the expert system are incorporated into the system design range and the practicability of the system is enhanced.

Keywords: dispatching and control integration; unattended centralized monitoring; equipment defect; data mining; pre-warning and pre-control

1 INTRODUCTION

In the operational mode of dispatching and control integration, all transformer substations are centrally unattended and monitored. The influence of equipment condition changes on the dispatching and control operation requires more meticulous control. It goes without saying that equipment defects are important for the judgment of equipment conditions. Information like defects and the influencing degree of defects should be deeply and comprehensively mastered and information like historical defects of important equipment should be summarized, analyzed and utilized as well, so as to assist the operation of dispatching and control integration. At present, the condition-based maintenance of equipment operation data is to anticipate the operation tendency and the defect risk of equipment according to the state evaluation standard. However, the method of summarizing occurrence rules of defects based on defect statistics completely depends on empirical values and lacks sufficient theoretical guidance. Defect analysis and condition-based maintenance are deeply integrated in this research project. The SVM model of the equipment operation is obtained through the training of the historical data of the power grid operation. The risk of defects can be confirmed and the accuracy of risk anticipation can be effectively improved in accordance with the current status information.

2 MODEL TRAINING

Original data of equipment operation is heterogeneously stored in multiple systems. It is obliged to carry out isomorphic and quantitative processing on current data in the earlier stage of model training. The basic flow of the training is firstly to classify, sort and quantify historical data and then to obtain the data model with SVM. The occurrence cycle of equipment defects can be calculated according to the trained model and the scientific maintenance or inspection cycle can be thus calculated.

Data preparation:

At present, accounting information, operation data and defect data of equipment are respectively stored in different systems. The first step of data mining analysis is to break through the system barrier and store necessary data in the database of the system.

Data processing:

Classification: classify defects in accordance with types of equipment.

Aggregation: Each type of equipment has accumulated a great deal of defective data. Different types of equipment should be classified according to types of defects in order to ensure the accuracy of defect analysis. Taking a transformer for example, there are mainly seven major reasons for defects: insulation aging, insulating oil degradation, malfunctions caused by overvoltage, casing destruction, lead insulation fault, magnetic-circuit fault and tap switch fault.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License 4.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

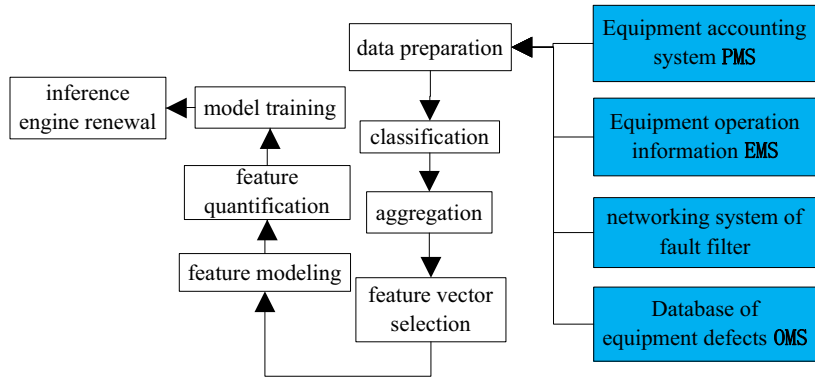


Figure 1. Flow chart of the training of SVM

Table1. List of defect vectors

Fault types	Features (fault causes)
Insulation aging	run time, temperature, humidity, load, voltage class, manufacture
Insulating oil degradation	run time, oil temperature, humidity, manufacture
Malfunctions caused by Overvoltage	run time, voltage, manufacture
Casing destruction	run time, temperature, humidity, material, manufacture
Lead insulation fault	run time, temperature, humidity, material, weather, manufacture
Magnetic-circuit fault	run time, temperature, humidity, material, weather, manufacture
Tap switch fault	run time, voltage class, temperature, weather, manufacture

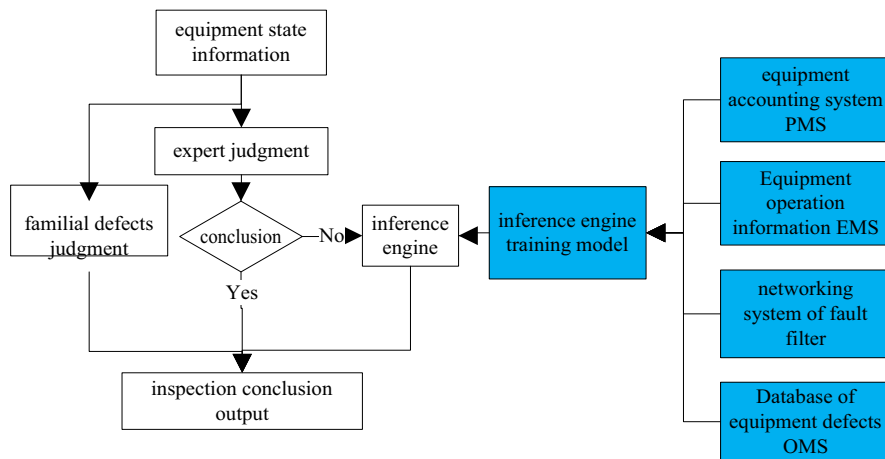


Figure 2. Overall design architecture of the system

Structures of defect database are various and description specifications of defects are different, so it is necessary to identify defect descriptions with advanced operation methods. There is now a mature and widely-used method based on decision-making tree and clustering, which divides transformer defects into 7 clusters through the methodology application. De-

fect data, which is not successfully divided into clusters, is abandoned as waste data.

Feature vector selection: Different factors for defects should be determined according to various types of defects, namely features. Feature vectors of the transformer in this research are selected in the table 1.

Feature modeling: A standard database should be

created for each defect record after the feature selection so as to establish a standardized data structure for subsequent data mining.

Feature quantification or dualization: The attribute of defects is text description, so it is necessary to carry out transformations on variables and defects should be transformed into required data types. For example, run time, temperature and humidity can be transformed into standard consecutive numbers. Weather conditions like rainfall can be differentiated with 0 and 1.

The process of the model training is as follows:

1) Select equipment and defect types to be trained.

Divide tested feature groups according into feature vectors of defects. The basis of selecting tested groups is that the feature group has sufficient data for statistical analysis on the premise that other features are operated in a safe region. For example, features of the insulation aging of a transformer include run time, temperature, humidity, load and voltage class.

2) Establish a model of SVM

$$y(X) = w^T X + b \quad (1)$$

Where, X_k is the input feature vector $\{x_1, x_2, \dots, x_k\}^T$, $w = \{w_1, w_2, \dots, w_k\}^T$ is the weight coefficient of feature vectors, and b is the offset value.

3) Model parameters identification

The objective function of the least squares SVM can be described as follows:

$$\min J(w, e) = \frac{1}{2} w^T w \quad (2)$$

The constraint condition:

$$y(X_i) = w^T X_i + b_i, i = 1, 2, \dots, n \quad (3)$$

The Lagrangian function is defined as follows:

$$L(w, b, \alpha) = J(w) - \sum_{i=1}^n \alpha_i \{w^T X_i + b - y_i\} \quad (4)$$

Where, α_i is the Lagrangian multiplier. According to the Karush-Kuhn-Tucker (KKT) optimum condition, the linear equations system of seeking partial derivatives of w and b are obtained as follows:

$$\begin{aligned} \frac{\partial L}{\partial w} = 0 &\Rightarrow w = \sum_{i=1}^n \alpha_i X_i \\ \frac{\partial L}{\partial b} = 0 &\Rightarrow \sum_{i=1}^n \alpha_i = 0 \end{aligned} \quad (5)$$

Substitute the above results into formula (4), so that:

$$\begin{aligned} L(w, b, \alpha) &= \frac{1}{2} w^T w - \sum_{i=1}^n \alpha_i \{w^T X_i + b - y_i\} \\ &= \frac{1}{2} w^T \left(\sum_{i=1}^n \alpha_i X_i \right) - w^T \sum_{i=1}^n \alpha_i X_i - b \sum_{i=1}^n \alpha_i + \sum_{i=1}^n \alpha_i y_i \quad (6) \\ &= \sum_{i=1}^n \alpha_i y_i - \frac{1}{2} \left(\sum_{i=1}^n \alpha_i X_i \right)^T \left(\sum_{i=1}^n \alpha_i X_i \right) \\ &= \sum_{i=1}^n \alpha_i y_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j X_i^T X_j \end{aligned}$$

It can be seen from formula (6) that the Lagrangian function value contains a variable α that can be solved with the Sequential Minimal Optimization (SMO) algorithm. After the solution of α , the model of SVM can be then obtained through the solution of w and b .

3 SYSTEM ARCHITECTURE DESIGN

The system provides three kinds of defect judgment mechanisms: familial defects judgment, expert judgment, support vector machine judgment. Judgment principles of familial defects judgment and expert judgment are made by users themselves, which are used to inspect whether there are defects or defect risks in the current operation status of equipment. The support vector machine judgment automatically trains the inference engine model according to the parameters preset by users. The reasoning accuracy will be continually improved with the increase of the training data. The support vector machine judgment is also a key point of the research project.

3.1 Inference engine judgment

The model of SVM is obtained through the training of power grid historical data and the probability of equipment operation in a future period of time can be then obtained by inputting the current operational variables of equipment.

The application process of the model:

1) Divide feature groups according to data regions. The basis for the division includes: safe region, high-risk region and out-of-limit region.

$$(a, b) \begin{cases} a < a_1, b < b_1 \\ a_1 < a < a_2, b_1 < b < b_2 \\ a > a_2, b > b_2 \end{cases} \quad (7)$$

2) Establish the incidence relationship between each data region and run time. In the experiment, the calculation starts from the lower limit of the high-risk region. The trend analysis on the probability of defects is carried out with the SVM model so as to obtain the basic rule of probability distribution. The histogram of defect probability is measured in week.

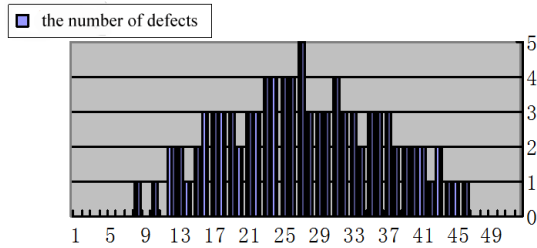


Figure 3. Distribution diagram of equipment defect data

3) Determine the safe region of the run time for each data region. Maintenance and inspection suggestions can be provided when the probability is larger than the threshold.

The safe operation time of each feature group in a certain region can be calculated through the above-mentioned 3 steps. When the equipment operation time exceeds the range, the system will provide maintenance and inspection suggestions of the equipment. Maintenance suggestions become more and more accurate with the increase of the training data.

The explanation of inference conclusions:

Model prediction results of SVM are directly shown to operators who are unable to interpret the meanings. Results should be “translated” into the natural language that can be understood by people so that users can comprehend the inference results and meanings correctly. Both the inference process and results are discrete values, so the formal language can be only formed through the translation of the interpreter with the addition of corresponding decorative texts. The interpreter is composed of a fact-driven unit, a rule translation unit and a display unit. The translation process is shown in the following figure:

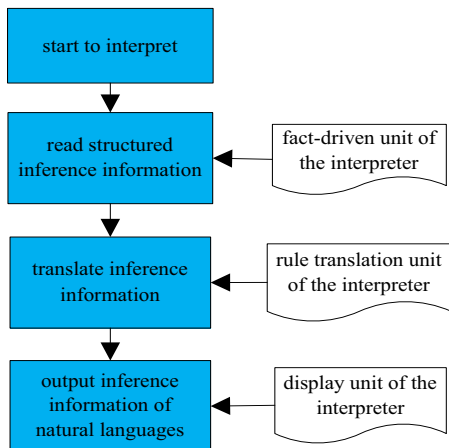


Figure 4. Process of outputting interpreted results

Advanced applications of the inference engine:

The inference engine is able to infer the probability

of equipment defects in a certain period of time in the future. But it is not the only evaluation index of equipment performance. Voltage, load, environment temperature and other factors have great influence on equipment performance. Excessive load and poor working condition accelerate the degradation of equipment performance and increase the probability of equipment failure tremendously. For this reason, a real-time status evaluation model for monitoring equipment should be constructed. The principle of the online evaluation model of real-time status is: parameters that can be measured online, including operation state parameters of power grid, operation state parameters of equipment and operation environment parameters of equipment, and the fault probability of equipment in off-working status are integrated together according to the corresponding integration rules so as to obtain an evaluated value of equipment operation state, which is called the index of operation capability. The operational logic of the real-time status evaluation of monitoring equipment is presented in Figure 5.

3.2 Expert judgment

Condition-based maintenance is a popular and effective equipment maintenance management method, which has played an irreplaceable role in system state maintenance. The basis of condition-based maintenance is that judgments and predictions of defects are provided by analyzing the current operation state data of equipment and referring to knowledge and experience of the power industry in the expert database. The knowledge base of the system is in the mode of the production rule, which presents typical experience of domain experts with rules so as to form a knowledge base of abnormal warning analyses. Through the user-oriented rules maintenance way, the knowledge base provides a new approach to convert rich specialized knowledge and operation processing experience of domain experts like monitoring staff and maintenance specialists into structured inference models. Being equipped with perfect learning ability and flexible expansion capability, it conducts maintenance and structured storage for experts’ knowledge through the definition of the fault model, the definition of the signal grouping model, the definition of the logical inference model and the definition of the time window.

Now, the online monitoring data of the power system is quite comprehensive. The expert system can effectively determine the current operation state of equipment by setting the limit of equipment operation data and the data growth rate. Expert judgment is still the most important basis for independent equipment maintenance plans in the system.

3.3 Familial defects judgment

Familial defects judgment requires rigorous analyses on equipment defects to determine the scope of equipment defects. There are now several defects judgment methods adopted by electric power companies:

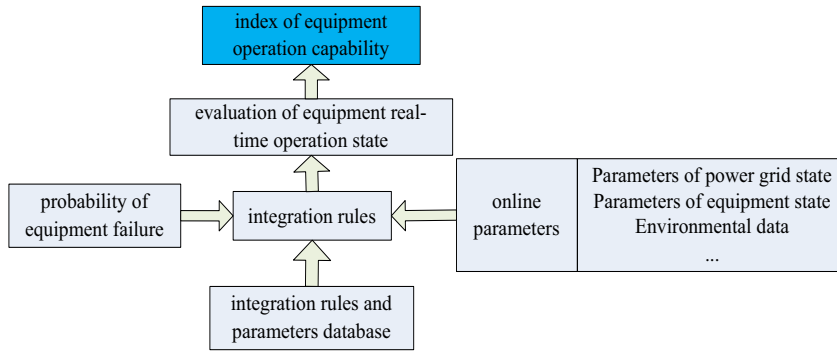


Figure 5. Process of equipment real-time evaluation

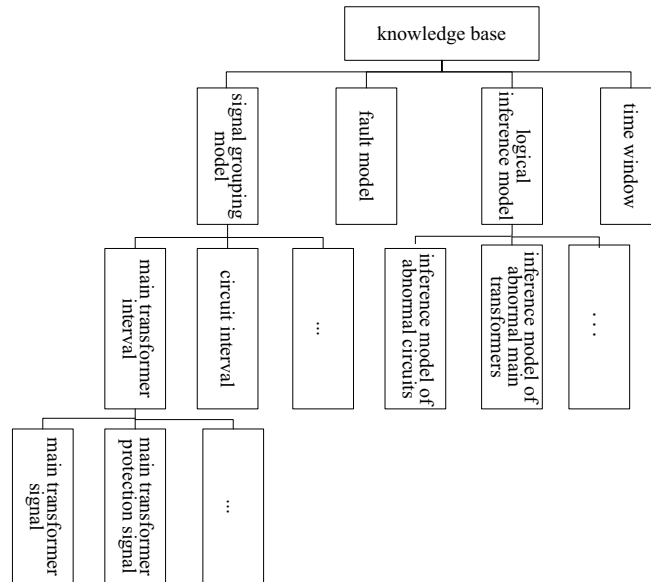


Figure 6. Structure of the knowledge base

- 1) Familial accidents notification of equipment transmitted by the superior;
- 2) Equipment countermeasures transmitted by the superior;
- 3) Familial defects of equipment transmitted by the superior;
- 4) Familial defects of equipment notified by manufacturers;
- 5) Familial defects of equipment provided by associate organizations.

The system also provides statistical data of familial defects judgment as the judgment basis so that all levels of electric power companies and plants can be equipped with the capability of conducting preliminary analyses on familial defects of equipment. The method is to provide statistics on whether there are obvious manufacturer features or material features in a

certain device with defects by making use of the data processed by inference trainings.

4 APPLICATION STATUS

State Grid Henan Electric Power Company has now successfully established a “defect management system of electric transmission and transformation equipment state monitoring” and realized the function of mining equipment maintenance rules from defective data. The system has successfully connected the equipment accounting information base in PMS, the equipment defects management base in OMS and the equipment operation state information base in EMS system. It has preliminarily possessed the ability of guiding and formulating maintenance and inspection plans through

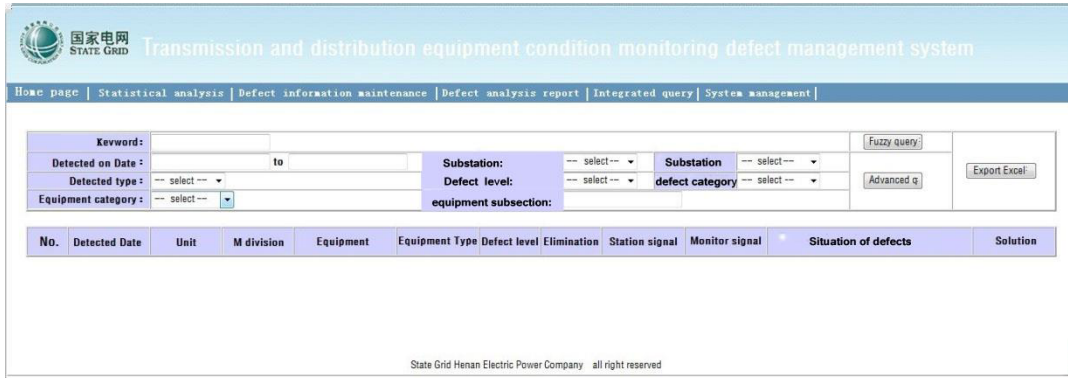


Figure 7. Electric equipment defects management system of Henan Province



Figure 8. Page of isomorphic defect information

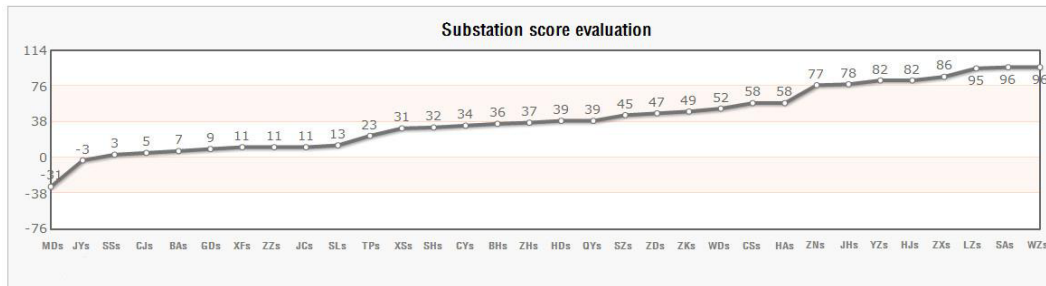


Figure 9. Operation ability indexes of a plant

data mining analyses.

The multidimensional query page will be automatically displayed after the login, in which such information as defect levels, descriptions and processing conditions can be overviewed. The figure is shown in Figure 7.

The system realizes data synchronization among the most central systems of dispatching and operation management, namely OMS, PMS and EMS, and completes the isomorphic storage of defect data. The query page of isomorphic data is shown in Figure 8.

The equipment operation ability index can be obtained by integrating operation state parameters of power grid, operation state parameters of equipment, operation environment parameters of equipment and the fault probability in a certain period of time in the future inferred by the inference engine. When the equipment operation ability index is smaller than the threshold, warnings will be provided to remind that

inspection and maintenance are needed. Meanwhile, the system integrates equipment operation ability indexes of a certain plant to obtain the operation ability index of the plant. The plant will be recommended to carry out overhaul when the score is lower than the threshold. The effect figure is shown in Figure 9.

5 CONCLUSION

This research project has been conducted in State Grid Henan Electric Power Company. The statistical function of the system provides familial defect judgment with an accurate data basis. In the meantime, the cycle of defect occurrence trained by the system is in line with practical situations, which verifies the effectiveness of the method. The inference engine finds out equipment operation rules from the big data of equipment defects and operation information. Compared

with a single device, the inference engine is of greater significance for the occurrence probability, influence degree and even processing methods of certain types of defects. The expert system of the condition-based maintenance theory has a greater advantage in defects diagnosis and prediction of independent equipment. The two are complementary in terms of advantages and can be integrally applied, which will play a more important role in the centralized monitoring and operational maintenance of substation equipment in the future.

ACKNOWLEDGEMENT

Fund Project: Natural Science Foundation of China (00000000); National High-tech R&D Program (863 Program) (2008AA000000).

REFERENCES

- [1] Margaret H Dunham. 2003. *Data Mining Introductory and Advanced Topics*. Upper Saddle River, N. J: Prentice Hall/ Pearson Education.
- [2] Soumen Chakrabartietal. 2009. Data mining: know it all. UK: Library of Congress Cataloging-in-Publication Data, 1~34.
- [3] Pan, Y.L. 2011. *A Study on the Improvement and Applications of Decision Tree Classification Algorithm (Master Dissertation)*. Kunming: Yunnan University of Finance and Economics.
- [4] Wang, F. 2013. *The Application of Decision Tree Algorithm in the Fault Diagnosis System of Mechanical Equipment (Master Dissertation)*. Wuhan: Huazhong University of Science and Technology.
- [5] Jia, Z. 2013. *A Study on the Application of Genetic Clustering Algorithm in Equipment Defects Classification (Master Dissertation)*. Shenyang: Liaoning Project Technology University.
- [6] Ma, J.H., Zhang, W.X. & Xu, Z.B. 2002. Data Mining and Database Knowledge Discovery: A View of Statistics. *Chinese Journal of Engineering Mathematics*, 19(1): 1-13.
- [7] Frank P. M. 1990. Fault Diagnosis in Dynamic Systems Using Analytical and Knowledge based Redundancy--A survey and Some New Results. *Automatica*, 26(3):459-474.
- [8] Wu, H.X., Zheng, X.J. & Pan, C.H. 1988. Numerical Solution and Automatic Measurement of Complex Dielectric Constants. *Journal of Microwaves*, 4(4): 22-29.
- [9] Zhang, Z.M., Xie, J. & Li, C.H. 2009. The Application of Data Mining in Defects Prediction and Management of Substation Equipment. *Journal of Harbin Institute of Technology*, 41(1): 272-274.