

# A Real-Time Intelligent Abnormity Diagnosis Platform in Electric Power System

Feng ZHAO\*, Guannan WANG\*\*, Chunyu DENG\*\*, Yue ZHAO\*\*

\*China Electric Power Research Institute, No. 15, Qinghe Xiaoying Road, Beijing, China

\*\* Department of Big Data Business, Beijing Guodiantong Corporation, No.1 Hangfeng Road, Beijing, China  
fengzhao@sgcc.com.cn, wgn1103@gmail.com, dengchunyu@sgepri.sgcc.com.cn, zhaoyue2@sgepri.sgcc.com.cn

**Abstract**—With the rapid development of smart grid, intelligent electric meters can be seen in most of the households, and the volume of electric energy data is in a rapid growth. This paper mainly aims at introducing an abnormity diagnosis platform in electric power system. It is used to distinguish the abnormal point according to the historical data and expert experience, and put forward some resolving scheme to ensure the high reliability and stability of power grid. In our approach, we use distributed technologies to process big electric energy data. Specifically, distributed file system (HDFS) and distributed database (HBase) are applied to data storage, and distributed computing technology (MapReduce) is applied to constructing knowledge base and computing. In the inference engine, we use Hidden Semi-Markov Model. This model can auto-get and modify knowledge in knowledge base, achieve a better real time phenomenon, through self-learning function and machine as well as interacting between human. The results show that this abnormity intelligent diagnoses platform is effective and faster.

**Keywords**— Abnormity Intelligent Diagnosis, Distributed Storage, Distributed Computing, Hidden Markov Model

## I. INTRODUCTION

At present, it has entered the digital, information age in the whole world. The advancements of information and communication technology (ICT) have enhanced our capability of collecting data, so does the electric power industry in China. For example, the intelligent electric meters upload electro-data every 15 minutes, then the number of electro-data in one day is  $96 \times N$ , where  $N$  is the number of electric power users which is usually more than 700million.

The big data in electric power system can bring us big opportunities. On one hand, the development of electric power industry and society are closely linked, effective application of big data in electric power system is the impetus of rapid development of society. On the other hand, a new study from University of Texas ([1]) shows that a minimal effort to improve data usability can have a massive impact on revenue. If there is a 10% improvement in data quality, electric service will have the largest improvement of ROE (Rate of Return on Common Stockholder's Equity). In other words, effective use of big data in electric power system can promote operation efficiency and profit ability dramatically.

Along with the rapid development of smart grid, the amount of data in electric industry is increasing too. Take only one

province as example, we have known that there is a big electro-data, even the number of abnormal information of consumer equipment in one day is more than ten millions. It is difficult to storage and compute these big data with traditional technologies.

In this paper, we mainly aims at analysing the large amounts of abnormal electro-data and introducing an intelligent abnormity diagnosis platform applied in electric power system. The abnormal electro-data includes the abnormal information of metrical equipment, acquisition equipment and electric power consumption of customer. This platform can distinguish the abnormal point according to the history data and expert experience, and put forward some resolving scheme to ensure the high reliability and stability of power network, therefore provide the directive function to operation and maintenance staffs.

The structure of this paper is as follows: Section II introduces the whole architecture and the advantages of distributed technologies (including distributed storage and computing technology) applied in the intelligent abnormity diagnosis platform. Section III applies expert system to diagnose the abnormal information of consumer equipment, the knowledge base and inference engine are emphatically introduced. Section IV presents the results of experiments and the running state of platform. Finally, we provide a conclusion and offer an outlook of future work in Section V.

## II. PLATFORM ARCHITECTURE

There is a rapid development of ICT and electric power industry, the combination of both two technologies makes us be faced with the challenge of explosive growth of big data in electric power system, especially the electro-data and abnormal information of electric equipment.

Being aims at solving the problem caused by the big data, we give some analysis, and apply distributed technologies to construct the platform. This section introduces the technology architecture of intelligent abnormity diagnosis platform.

### A. Technology Architecture

Figure 1 shows the technology architecture of intelligent abnormity diagnosis platform. The detailed description of each module are as follows:

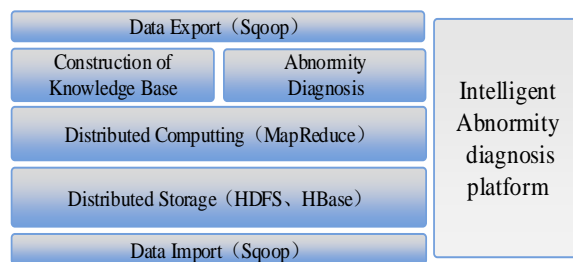
**Data import:** The main function of this module is data transforming, from relational database (Oracle). Sqoop is used

in this model, as it is a powerful Extraction-Transformation-Loading tool, especially at data transforming between comprehensive database and knowledge.

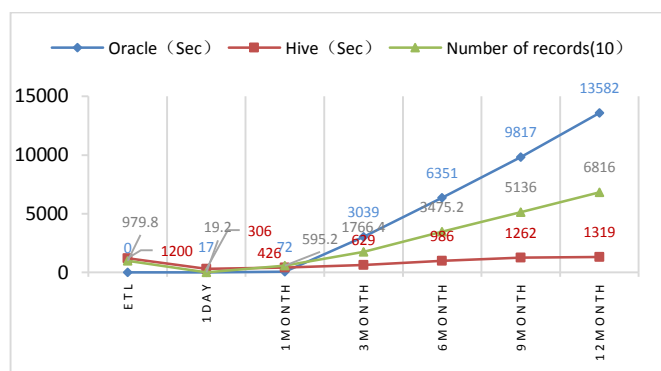
**Storage:** The main function of this module is storing and updating the knowledge from relational database and experts. Distributed file system (HDFS) and distributed storage (HBase) are used in this module.

**Computing:** This module executes algorithms to allocate the reason for abnormalities according to the knowledges. Distributed computing technology (Mapreduce) is applied.

**Data export:** This module transfers diagnosis results to the database, and make a backup in HBase. Sqoop is also hired here.



**Figure 1.** The technology architecture of intelligent abnormality diagnosis platform



**Figure 2.** The comparison of efficiency of distributed technology and traditional technology

In order to compare the efficiency of distributed technology and traditional technology, we compute the statistical characters (mean, variance) of abnormal case by using Oracle and Hive (a data warehouse infrastructure built on top of Hadoop). Figure 2 shows the elapse time of two technologies. The data was collected from a county in China. Figure 2 suggests following conclusion:

1) Comparing to traditional technologies, distributed technologies increase computing speed observably, especially as data gets bigger.

2) The computing efficiency of distributed technology and traditional technology are not significantly different over small data.

### III.EXPERT SYSTEM

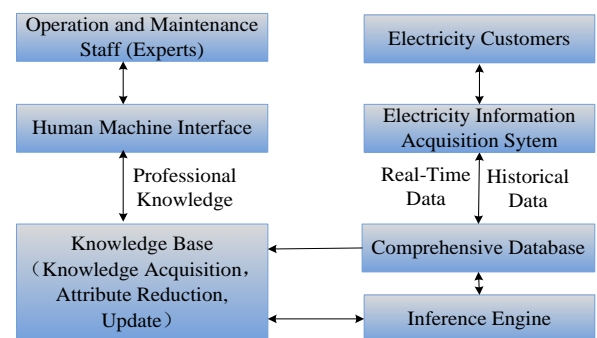
Aiming at distinguishing the abnormal point of the consumer equipment and consumer, expert system is applied to do this work. This section mainly introduces the basic working principle of expert system, knowledge base and inference engine.

#### A. Basic Principle

The expert system is applied to distinguish the abnormal point according to the history abnormal electro-data and expert experience. In the intelligent abnormality diagnosis platform, expert system is composed of four parts: human machine interface, comprehensive database, knowledge base and inference engine, and the last two parts are the most important parts in the system. Figure 3 shows the structure graph of expert system. The arrows represent the direction of data flow.

In the platform, an abnormal case is composed of some abnormal phenomenon and a corresponding reason of such phenomenon. The abnormal phenomenon are obtained from electricity information acquisition system, and the reason of abnormality (professional knowledge/experience of experts) is obtained from experts. Experts are the operation and maintenance staffs, and they are also the users of platform.

The functions of each part in the platform are given as follows:



**Figure 3.** The structure graph of expert system in intelligent abnormal diagnosis platform

1) Human machine interface is applied to show the possible results of diagnosis and get the reason of abnormalities from the experts/users.

2) Comprehensive database is applied to storage the real-time data and historical data, including the occurrence time and location of abnormal case, the abnormal phenomenon, reason of abnormality, power curve, name of abnormal equipment and others.

3) Knowledge base is applied to storage the abnormal phenomenon of abnormal case and the reason of such phenomenon. Knowledge base is required to have four basic functions: knowledge acquisition, knowledge storage, attributes reduction and knowledge base update. Knowledge base can realize its value through inference engine.

4) Inference Engine is applied to compute the abnormal point according to the real-time data (abnormal phenomenon)

and the knowledge in knowledge base. The algorithm we used in inference engine is Hidden Semi- Markov Model.

TABLE 1. AN EXAMPLE OF DECISION TABLE

Number	Reason of Abnormity	Anomaly Type	Abnormal Phenomenon					Occurrence Number
			Phenomenon 1	Phenomenon 2	Phenomenon 3	Phenomenon 4	.....	
1	Reason 1	Metric Equipment	1	1	0	0	.....	$N_1$
2	Reason 2	Acquisition	1	1	0	0	.....	$N_2$
3	Reason 3	Metric Equipment	0	0	1	1	.....	$N_3$
4	Reason 4	Acquisition	0	0	1	1	.....	$N_4$
5	Reason 5	Abnormal in Using Electricity	0	1	0	0	.....	$N_5$
.....	.....	.....	.....	.....	.....	.....	.....	.....

Knowledge base and inference engine are the most important parts in the expert system.

### B. Knowledge Base

There are three kinds of anomaly types, including abnormal information of metric equipment, acquisition and electric power consumption of customer. TABLE 1 shows an example of decision table which includes Reason of Abnormities, anomaly types, and abnormal phenomena.

The functions of knowledge base are given as follows:

#### 1) Knowledge acquisition

In the platform, knowledge in the knowledge base means Reason of Abnormities and abnormal phenomena of abnormal case which are obtained from operation and maintenance staff and electricity information acquisition system respectively.

Sqoop is applied to transform the data (abnormal phenomenon) from comprehensive database to knowledge base.

#### 2) Knowledge storage

Decision table is the core of knowledge base. TABLE 1 shows an example of decision table. Take province Ningxia as example, by analyzing the abnormal electro-data, we find that there are  $N_r$  reason of abnormities,  $N_p$  abnormal phenomenon. Let  $R_i$  be the  $i$ th Reason of Abnormity,  $P_j$  be the  $j$ th abnormal phenomenon, where  $i = 1, 2, 3, \dots, N_r$  and  $j = 1, 2, 3, \dots, N_p$ .

All the abnormal case in decision table should apply with the rule as

$$C^l : P_1 + \dots + P_{N_p} \rightarrow R_i, \quad (1)$$

where  $C^l$  is the  $l$ th abnormal case in decision table. This rule has the following meaning: in the abnormal case  $C^l$ , reason  $R_i$  is the cause of phenomenon  $P_1, \dots, P_k$ . Take a real abnormal case as example. Let  $P_1$  is the phenomenon “no voltage curve in the terminal”,  $R_1$  is the corresponding reason “failure in data transfer”,  $R_2$  is the another corresponding reason “no voltage curve in the electric energy meter”. Actually,  $P_1$  can be caused by  $R_1$  or  $R_2$ . In one word, the same phenomenon can be caused by different reasons, conversely, the same reason can lead to different phenomenon.

We did not storage all the abnormal cases in the decision table, the occurrence number of each kind of abnormal case is listed in the last column. This kind of decision table can help the platform saving search time in the progress of attribute reduction and knowledge base update.

#### 3) Attribute reduction

Rough set theory is a new mathematical tool to deal with vagueness and uncertainty.

In the real data, some abnormal phenomenon tend to appear in the same time, so it is not necessary to list all the phenomenon in the decision table. Furthermore, by analyzing the characters of abnormal case, we can see that the initial decision table based on raw data is an inconsistent decision table, then the traditional reduction method is not suitable to analyse the abnormal data in electric system. So we apply a high efficient algorithm based on  $\lambda$ -reduction to solve this problem.

The work is conducted in 5 steps as follows:

Firstly, establish the initial decision table according to the raw data mentioned above.

Secondly, compute the decision vector  $\mu_B$ .

The decision table can also be defined as  $DT = \{U, B \cup D, f\}$ , where

$$U = (C_1, C_2, \dots, C_{N_c})$$

$$B = \{P_1, P_2, \dots, P_{N_p}\}$$

$$D = \{R_1, R_2, \dots, R_{N_r}\}$$

According to the initial decision table, it is easy to obtain  $U/B = \{B_1, B_2, \dots, B_n\}$  and  $U/D = \{D_1, D_2, \dots, D_m\}$ .  $\forall x \in U$ , the decision vector can be defined as

$$\mu_B = (D(\frac{D_1}{[x]_B}), \dots, D(\frac{D_m}{[x]_B})),$$

$$\text{where } D(\frac{D_i}{[x]_B}) = \frac{|D_i \cap [x]_B|}{|[x]_B|}.$$

Thirdly, compute the fuzzy equivalent relation matrix  $TS_\lambda$ .

$\forall x, y \in U$ ,  $\mu_B(x) = (\mu_{x_1}, \dots, \mu_{x_n})$ ,  $\mu_B(y) = (\mu_{y_1}, \dots, \mu_{y_n})$ , then the fuzzy equivalent relation matrix is defined as

$TS_\lambda = \{(x, y) \in U \times U \mid TS(x, y) \geq \lambda\}$ , where  $TS = S_k$ ,  $k \geq |U|$ , and

$$S(x, y) = \frac{\sum_{i=1}^n \mu x_i \times \mu y_i}{\sqrt{\sum_{i=1}^n \mu x_i^2} \times \sqrt{\sum_{i=1}^n \mu y_i^2}}$$

Fourthly, compute the  $\lambda$ -reduction.

$U/TS_\lambda$  can be obtained in the above three steps, then we can compute the  $\lambda$ -reduction by using the heuristic algorithm based on sorting proposed by [2].

We can see that  $\lambda$ -reduction is the best attribute reduction.

#### 4) Knowledge base Update

As time went by, the best attribute reduction is not invariable because abnormal cases are changing. The knowledge base needs to be able to add new abnormal case in time. If the new abnormal case exists in the decision table, add one to the occurrence time, and if the new abnormal case does not exist in the decision table, add a row, and the occurrence time is one.

### C. Inference Engine

The Algorithm of abnormality diagnosis is the core in inference engine. Hidden Semi-Markov Model (HSMM) is applied to diagnose the abnormal reason of equipment and abnormality in using electricity.

A HSMM is a statistical model with the same structure as semi-Markov rather than Markov. This means that the probability of there being a change in the hidden state depends on the amount of time that has elapsed since entre into the current state.

The parameter set of HSMM is denoted as  $\mu = (N_r, N_p, \pi, A, B, P)$ . Figure 3 shows the progress of diagnosis in inference engine in detail. We aims at finding the most suitable HSMM according to the known knowledge.

1)  $N_r$  is the number of states in the state set  $R$ ,  $R = \{R_1, R_2, \dots, R_{N_r}\}$ , it is the abnormal reason set in the paper.

2)  $N_p$  is the number of observation sequence (output sequence),  $P = \{P_1, P_2, \dots, P_{N_p}\}$ . We see the abnormal phenomenon as output sequence.

3)  $\pi$  is the initial state probability.  $\pi = \{\pi_1, \pi_2, \dots, \pi_{N_r}\}$ , where  $\pi_i$  is the probability when the initial state is  $i$ ,  $\pi_i := P(q_1 = i)$  with  $\sum \pi_i = 1$ .

4)  $A = [a_{ij}]_{N_r \times N_r}$  is the state transition probability matrix (STPM). Considering the first order Markov Chain, the current state  $q_i$  only depends on the previous state  $q_{i-1}$ , that is  $a_{ij} = P(q_i = S_j | q_{i-1} = S_i)$ .

5)  $B = [b_j(k)]_{N_p \times N_r}$  is the state output probability distribution. In our paper,  $\pi, A, B$  are unknown parameters.

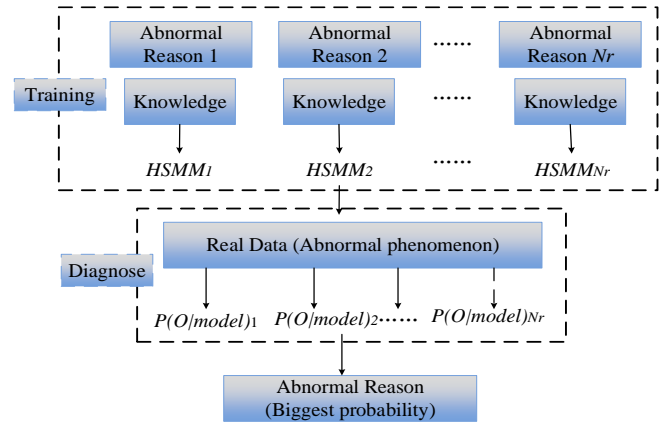


Figure 4. The progress of diagnosis in inference engine

6)  $P = \{P_1(d), P(d)_2, \dots, P_{ns}(d)\}$  is the set of state duration probability.  $P_i(d)$  is the probability of duration time the in state  $i$ ,

$$P_i(d) := P(q_{t+d+1} \neq j, q_{t+d} = j, \dots, q_{t+2} = j | q_{t+1} = j, q_t = i).$$

Assuming that the  $i$ th state duration distribution are given by Gaussian density function with mean  $\mu$  and variance  $\sigma$ .

$$P_i(d) = N(d | \mu, \sigma)$$

$\mu$  and  $\sigma$  are estimated with mean value  $\overline{pt}$  and  $\sum_{i=1}^{N_p} (pt_i - \overline{pt})^2 / (N_p - 1)$ .

Baum-Welch algorithm (also known as Forward-Backward algorithm) is used to find the most suitable HSMM.

### IV. EXPERIMENTS

In this section, we intend to show the efficiency of the abnormality intelligent diagnosis platform over some experiments. The data was collected from a county in China. Period of time is 1 month, 2 months, ..., and 12 months (1 year), obviously, the amounts of abnormal case increases as the period changes.

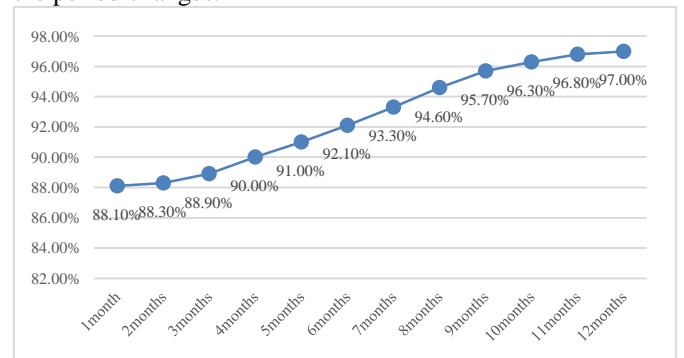


Figure 5. The result of abnormality intelligent diagnosis platform

Figure 5 shows the correctness of our platform, and the correctness is expressed by  $R_t = N_t / N$ , where  $R_t$  is the

correct rate,  $N_i$  is the number of diagnosis, and  $N$  is the number of abnormal case we need to diagnose. From Figure 5, we have the following conclusions:

1) The interval of correct rate is [88.1%, 97.3%], so the experiment results are satisfied;

2) With the increase in the amount of historical data, the correctness increase too. However, the rate of increase is not invariable. At first,  $R_i$  increase slowly, then there is a relatively rapid increase, at last, the speed of increase turns slower again.

## V. CONCLUSIONS

In this paper, we mainly aims at describing the abnormity intelligent diagnosis platform constructed in electric power system. In the platform, distributed technologies and tools are applied to solve the problem of big data storage and computing, expert system is applied to analyze the characteristic of big data and identify abnormal points according to the historical data and expert experience, and put forward some resolving scheme to ensure the high reliability and stability of power grid. Finally, the experiment shows a high efficiency and accuracy of the platform. In the power system, this is the first approach of constructing an abnormity intelligent diagnosis platform based on distributed technologies and expert system which helps in dealing with big data. Our future work focus more on abnormal electro-data analysis and set up an update mechanism of this platform.

## ACKNOWLEDGMENT

Many people have made invaluable contributions, both directly and indirectly to this paper. The engineer Xiaofeng Hu and other technicians offered help in the section of technology architecture. Siqi He and other data analysts offered help in analyzing the characteristic of data. I would like to express my gratitude to all those colleagues in department of Big Data who have offered valuable suggestions in making the paper a reality.

## REFERENCES

- [1] J. Bulla, I. Bulla, and O. Nenadić, "Hsmm—An R package for analyzing hidden semi-Markov models," *Computational Statistics & Data Analysis*, vol. 54, pp. 611-619. March 2010.
- [2] A. Barua, D. Mani, and R. Mukherjee, "Measuring the business impacts of effective data," Chapter One of a Three-Part Study, 2010.
- [3] S. Y. Zheng, "Hidden Semi-Markov Models," *Artificial Intelligence*, vol. 174, pp. 215-243. February 2009.
- [4] S. H. Liu, Q. J. Sheng, B. Gao, Z. Z. Shi, and F. Hu, "Research on efficiency algorithms for rough set methods," *Chinese Journal of Computers*, vol. 26, pp. 524-529. May 2003.
- [5] B. Han, T. J. Wu, and M. H. Yang, "Dynamic reduct research based on rough set theory," *Systems Engineering - Theory & Practice*, vol. 22, pp. 67-73. June 2002.
- [6] J. Du and Y. Ren, "An expert system model of malfunction intelligent diagnose," *Electro- Optic Technology Application*, vol. 24, pp. 73-76. August 2009.
- [7] M. C. An, "A Survey on fault diagnosis expert systems," *Computer Measure & Control*, vol. 16, pp. 1217-1219. September, 2008.
- [8] Y. T. Zhang, Z. Y. He, J. Zhao, P. Zhang, M. Li, and J. T. Gui, "A power network fault diagnosis method based on rough set theory and naïve Bayesian networks," *Power System Technology*, vol. 31, pp. 37-43. January 2007.

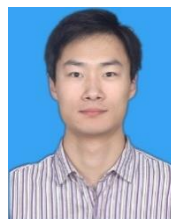
- [9] T. S. Bi, Y. X. Ni, and F. L. Wu, "Hybrid fault section estimation system with radial basis function neural network and fuzzy system," in *Proc. CSEE '05*, 2005, p. 64-48.
- [10] R. Chang and H. W. Wang, "An Algorithm for training parameters in belief rule-bases based on gradient methods and optimization step size," *Journal of North China Institute of Water Conservancy and Hydroelectric Power*. Vol. 1, pp. 154-157. February 2011.
- [11] C. Angeli, "Online expert systems for fault diagnosis in technical processes," *Expert Systems*, vol. 25, pp. 115-132. May, 2008.
- [12] C. Angeli and A. Chatziniikolaou, "On-line detection techniques for technical systems: A survey," *International Journal of Computer Science & Applications*, vol. 1, pp. 12-30. January, 2004.
- [13] G. Xiang, H. B. Zhang, H. M. Li, and X. Y. Chen, "The research of electronics equipment online failure prediction based on gray theory and expert system," *Aerospace Control*, vol. 31, pp. 21-26. September, 2013.
- [14] Q. Miao, "Condition monitoring of rotating machinery using hidden markov models," *Astonautica Sinica*, vol. 26, pp. 641-646. September, 2005.
- [15] Z. K. Peng and F. L. Chu, "Application of wavelet transform in machine condition monitoring and fault diagnosis: a review with bibliography," *Mechanical System and Signal Processing*, vol. 18, pp. 199-221. March, 2004.



Feng Zhao was born in HeBei province, China, October, 14th, 1980. He was graduated from Beijing University of Posts and Telecommunications, master, majoring in automation and now he is a PHD student in China Electric Power Research Institute., majoring in Electrical Engineering and Automation.



Guannan Wang was born in Hebei China, November, 3rd, 1985. She received her Doctor's degree in mathematical statistics from the Central South University in 2013. Dr. Wang is devoted to the work of data analysis in the power industry.



Chunyu Deng was born in Hebei China, 1982. He was graduated from North China Electric Power University, master, majoring in computer science and technology. Mr. Deng is devoted to the work of application of data mining in power industry.



Yue Zhao was born in Jiangsu China on July 16th, 1973. She received her Master's degree in Science from the University of Massachusetts Amerhest in 2002 and worked as a statistical analyst in Boston University between 2002 and 2008. Yue Zhao joined Beijing Guodiantong Corporation in 2008 and has been devoted to electric data analysis since then.