# Research on Condition Monitoring of Power Big Data Based on Rough Sets

Yulong Yan[1, a], Jilai Wu [2], Shejun Wu [2] and Jian Zhang[1, b]

[1]School of Electrical Engineering, Zhengzhou University, Zhengzhou 450000, China

[2] State Grid Ningbo Yinzhou Power Supply Company, Ningbo 315000, China

[a]yylong0280@163.com, [b]zhangjian63@zzu.edu.cn

**Keywords:** power big data, data mining, condition monitoring, rough set.

**Abstract.** The urgency of demanding for large data in power industry will greatly exceed that of in other basic energy industries in the future. Through analyzing the power equipment condition monitoring data, we can assess more accurately the operating condition of power equipment, predict the life span of the equipment, and thus can effectively avoid economic losses caused by excessive supervision as well as shortness of supervision. To further promote the wide and deep application of power big data technology in condition monitoring, this paper summarizes equipment condition monitoring studies based on big data technology, and discusses applications of the key technologies of Big Data-Data Mining in condition monitoring. For data mining techniques, this paper uses a rough set theory to study application in knowledge and data discovery and fault monitoring. And for the presence of some problems of rough set theory applied in practical systems, some directions for further research are proposed.

## 1. Introduction

With increasing development of the breadth and depth of power transmission equipment, more and more monitoring data has been collected, consequently accumulating a large smart grid condition monitoring data. However, how to store and analyze effectively condition monitoring large data is one of the key issues of large data applied in the field of condition monitoring. On the basis of the characteristics of state monitoring data and cloud-based platform, it is possible to combine massive of small files monitoring data into a large sequence of files and choose the way of compressed storage, thus improving its efficiency of storage and processing [1].

Looking into the current state of development and application of grid monitoring technologies, condition monitoring technology is a technology for state-based monitoring or predictive maintenance services, and its development is derived from state supervision status for the technical demand of power grid equipment acquisition of state information, analysis, and evaluation. Therefore, at present, the development of condition monitoring technology is aimed at one particular type of equipment or a kind of concrete application, mainly concentrating on the power generation equipment, substation facilities, transmission line and other specific equipment supervisor applications [2].

Big data typically has a "3V" feature, meaning large volume, variety and low density value [3]. And the problem of the low density value in condition monitoring big data is particularly thorny. Most of the monitoring data represents the normal operation of equipment, but few is valuable and can effectively reflect the characteristic of the equipment operation condition. Features data is analyzed and sample data is dynamically enriched and updated, which contributes to the search of valuable feature from condition monitoring big data. Transmission equipment condition monitoring system is based on big data, capable of solving inherent problems that exist in electrical equipment condition monitoring and diagnosis, which serves as an important means of achieving state power transmission equipment supervision management and improving lean level of power production operation and management.

 

## 2. Equipment condition monitoring technology based on big data

The basic processing of big data is not much different from the traditional data processing, while that big data needs to deal with huge amount of unstructured data is a big different between the two processes. As a result, parallel processing should be adopted in every link so as to meet the demand for mass data processing.

### 2.1 Power Equipment overall technical program

A common power equipment integrated monitoring system architecture based on big data is shown in Figure 1.

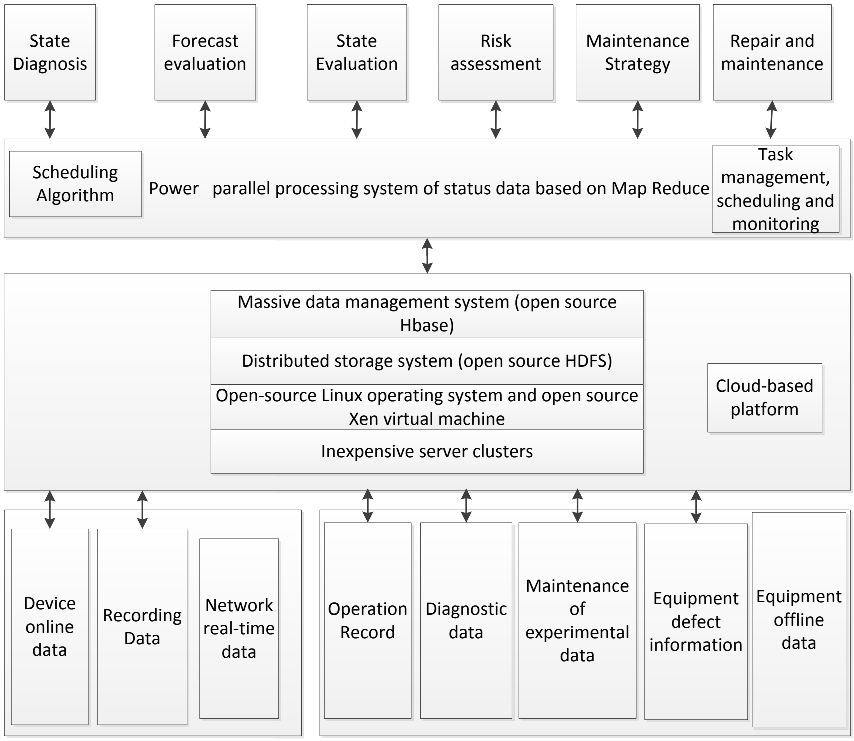


Fig.1 Power equipment of big data monitoring system architecture diagram [4]

The data sources of integrated monitoring system contains grid real-time data and equipment offline status data. Network real-time data includes not only the equipment on-line condition monitoring data, but also the power grid running conditions data, fault recording data, etc. Equipment offline data contains equipment operation records, diagnostic data, supervision of experimental data, the device defect information and accounting of equipment archives. Power grid real-time data and equipment offline data are centrally managed through cloud-based platform. The cloud-based platform is made up of inexpensive server clusters based on open source of HDFS distributed storage system and open source HBASE of mass data management systems and other components[5]. The power grid state data parallel processing system based on MapReduce above the cloud file storage platform contains an algorithm scheduling, task management, scheduling, monitoring and other functional modules. Based on cloud file storage platform and grid status data parallel processing system, it is possible to carry out advanced applications based on big data technologies, including the status of diagnostic, prognostic assessment, condition evaluation, risk assessment, supervision policy, supervision and maintenance functions etc.

On the basis of power equipment integrated monitoring system structure diagram, it is easy to obtain the overall technical solution diagram of power equipment monitoring based on big data. Figure 2 is one kind of the overall program diagrams. The relevant data is obtained via monitoring cables and cable accessories, switchgear monitoring, transformer monitoring, GIS monitoring, but laboratory test data and expert knowledge are acquired through a series methods such as data mining, characterization and knowledge status assessment to finally determine the equipment status levels. With overall monitoring system based on big data platform, equipment monitoring data is analyzed and dealt with, and then predictions like the life span of equipment or the suggestion for equipment maintenance can be drawn using state diagnostic model.

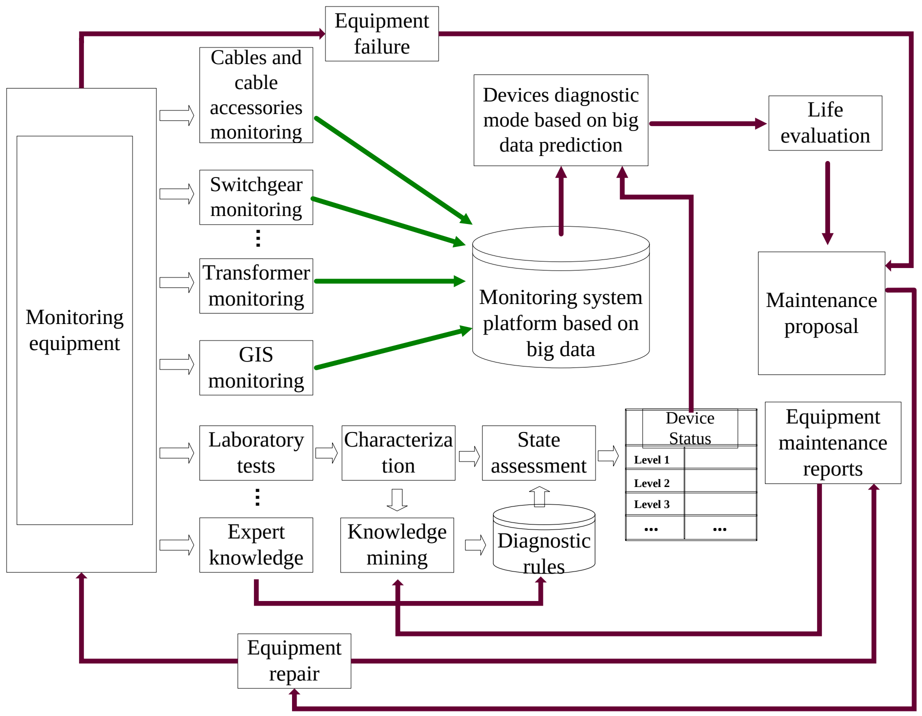


Fig.2 Power equipment based on big data integrated monitoring system

## 2.2 Big Data Processing.

The amount of state data in smart grid will surge significantly, which is far beyond the ability of traditional power grid condition monitoring, including not only the real-time online data, but also basic equipment information, test data, operating data, defect data, patrol supervisor records, live test data and other offline data. The requirement of reliability and instantaneity is extremely urgent. Faced with these massive, distributed, heterogeneous and complex real-time status data, it's difficult for conventional memory processing technology to meet the requirements, although MapReduce technology can provide support for large-scale data of parallel processing through a simple programming interface. However, existing MapReduce methods such as Hadoop, Phoenix, etc., belongs to batch mode for the persistent data [6]. During every process, the operating environment is needed to be initialized, and large-scale data is loaded and dealt repeatedly. Map and Reduce processes are also needed to be executed synchronously and finally large amounts of data between nodes is transmitted. Continuous arrival flow of data is dealt with in batch mode. If every time the small batch of data is processed, we will face the following problems: the expenditure of this system is too high, and instantaneity is also limited. If the waiting batch reaches a certain size, it will also increases the processing delay, as a result it still cannot meet the real-time demand [7].

In order to solve the problems caused by processing of real-time massive data, at present there are three ways out as follows [8]:

1. Stream processing mode. Data stream processing technology has developed rapidly over the past few decades, and is widely used in network traffic monitoring, the field of financial information processing, Web applications, and sensor networks. Stream technology is mainly used for online analytical processing, so the demand for time performance of algorithm is high. Data stream processing technology can provide effective support for the real-time processing and analysis of power information, as well as taking reasonable decisions to reduce the possibility of systemic risk. Although stream processing mode is naturally suitable for real-time processing system, the fields of its application is relatively limited.

2. Batch mode. In recent years, the use of batch mode to develop real-time system has become a research hotspot and achieve a lot of attainments. From the perspective of incremental computing, Google has proposed incremental processing system Percolator, and Microsoft has put forward Nectar and Dryadinc. The three are all available to incremental computing of large-scale data, which is very helpful for improving the real-time processing ability of massive data. But the disadvantage is that these systems are not compatible with MapReduce.

3. The combination of stream processing and batch mode. Many researchers have tried to merge stream processing with batch mode, and the main idea is to achieve the aim of stream processing using MapReduce model. For example, C-MR system achieves the continuous MapReduce (Continuous-MapReduce) of supporting streaming through three aspects of the work: (1) applying window concept of the parallel stream processing to MapReduce model; (2) Effective combination of a variety of heterogeneous computing capabilities including CPU and GPU; (3) Supporting flexible and dynamic workflow schedulability.

## 3. Power big data based on rough sets

### 3.1 Rough set reduction of condition monitoring big data

Combining the computing capability of cloud platform with the effectiveness of the greedy algorithm based on rough set theory, parallel algorithms based on MapReduce are designed. Big data samples are simplified so as to obtain plenty of simplifications equivalent to original information. Decision rules are also extracted to reveal the redundancy of fault information collection, consequently solving the problem of massive data reduction. The basic idea is that every time the greatest attribute or attribute combination which has the largest role in classification is selected as a candidate attribute put into the reduction collection. From the initial state of problem, optimum value is obtained with a number of greedy choices. The successive greedy choice is mainly made with top-down and iteration method. Every time once a greedy choice is selected, the original problem is simplified to a smaller scale similar to sub-problems.

### 3.2 Data mining algorithm based on rough set theory

There are three main stages in data mining: data preparation, data mining, evaluation and presentation of results. The main work in the first stage is the completion of large amounts of data selection, purification, speculation, conversion, data reduction, which will directly affect the efficiency and accuracy of data mining as well as the effectiveness of the final model. In the second stage, at first we need to select appropriate mining algorithms, such as decision trees, classification, clustering, rough sets, association rules, neural networks, genetic algorithms and so on, then analyze the data to obtain the knowledge model. The evaluation and expression of results is mainly to determine the validity of knowledge pattern models in order to find a meaningful model.

As for electrical equipment on-line monitoring system based on big data, it is significant to find parameters associated with electrical equipment by the means of data mining technique in order to gain power equipment status assessment.

Rough set method first discretizes attribute of information system with similar methods, then divides each property into equivalence classes, and simplifies information systems with the equivalence relations of set. Finally a minimal decision relationship can be acquired, as well as rules [9]. Currently mature relational database management system and newly developed data warehouse management system have laid a solid foundation for data mining based on rough set [10].

To take the distribution network fault diagnosis based on rough set theory as an example. The Decision Table Reduction of RS theory is applied to distribution network fault diagnosis, and the signal of protection and circuit breakers is considered as a condition attribute set of fault classification. Finally the decision table is established after considering various fault conditions that may occur. First original information is simplified with the help of the strong simplification capability of rough set theory so as to obtain lots of simplifications that are equivalent to the original information, and after further simplification, the smallest simplification can be found, and decision rules are chose to reveal the redundancy of fault information collection. $DS_0$ is defined as the decision attribute of the original decision table. The steps of simplifying decision table are as follows [11]:

(1)Finding all simplifications. Recursive call is employed in this paper to achieve the simplification of N condition attributes. When the k-th condition attribute is deleted, decision attribute is denoted by $DS_k$ . If $DS_k$ is different from $DS_0$ , then it indicates that the condition attribute can't be omitted. On the contrary, the condition attribute can be omitted, and the remaining N-1 condition attributes are further simplified.

(2) After further simplification the smallest reduction can be found. It is assumed that each simplification contains L decision rules and R condition attributes. First the j-th condition attribute $a_{ij}$ of i-th decision rule is deleted, then the remainder R-1 condition attributes $a_{in}$ $(n = 1 \sim R , n \neq j)$ of the decision rule is compared with condition attribute $a_{mn}$ (m = 1 ~ L, m ≠ i) of the rest of L-1 decision rules. If the conditions are identical, then whether these decision attributes are the same or not is needed to be analyzed. If they are identical, it means that the deleted condition attributes $a_{ij}$ is redundant, and on the contrary it is not redundant. The flow diagram judging whether a condition attribute $a_{ij}$ is redundant is shown in Figure 3, and so on, we achieve a minimum reduction.

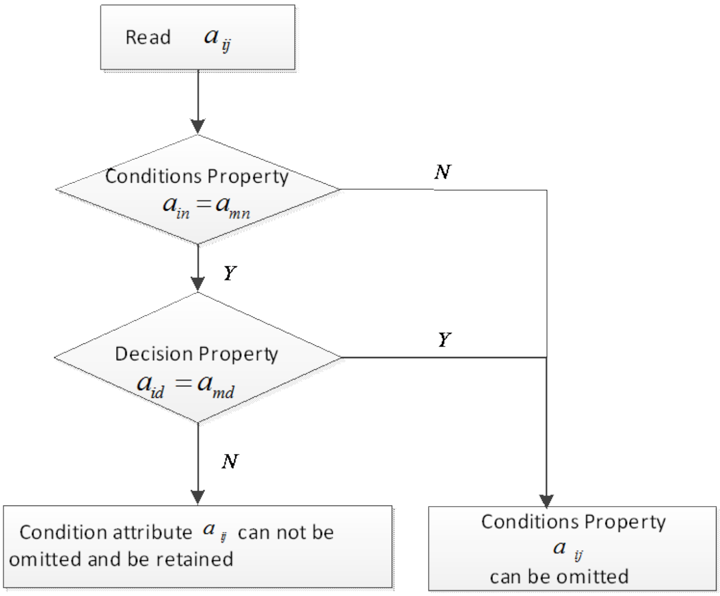


Fig.3 Find the most simple reduction flowchart

## 4. Conclusions

Compared with the traditional power grid, electrical equipment condition monitoring system based on big data can achieve the aim of equipment intelligence and informationization by installing monitoring devices based on advanced sensor technology, consequently achieving the status of self-aware device and fault self-diagnosis. It provides required data of decision-making for the production management, monitoring equipment selection and smart grid decision analysis technology. The basic principle is that through a variety of sensor technology, wide-area communications technology and information processing technology, we can achieve real-time sensing, analysis and diagnosis, prediction and evaluation of power equipment operating status, and its construction and extension work is also meaningful to improve the level of smart power grid.

Rough set theory is an effective method of handling incomplete information. It has a strong qualitative analytical skill, and it's helpful to find out the inherent law, but there are also complex query rules and other issues. To solve these problems, rough set theory is introduced into fault diagnosis system of generators, cables and transformer [12]. By applying the rough set method, it's possible to fuse multi-source data and realize establish and effective maintenance of complete knowledge base for fault diagnosis expert system.

In the future through analyzing the power equipment condition monitoring data, we can assess more accurately the operating condition of power equipment, and power big data technology with rough set will be widely applied to the field of condition monitoring.

## References


[1] Zhou Guoliang, Song Yaqi, Wan Guilan, Zhu Shuili. Research of condition monitoring big data storage and clustering. Transactions of China Electrotechnical Society. Vol. 28 (2013) No. 2, p. 337-344.

[2] Liu Ji, Huang Guofang, Xu Shiming. Development of smart grid condition monitoring. Electric Power Construction. Vol. 30 (2009) No. 7, p. 337-344.

[3] Song Yaqi, Zhou Guoliang, Zhu Shuili. Present status and challenges of big data processing in smart grid. Power System Technology. Vol. 37 (2013) No. 2, p. 927-935.

[4] Wang Dewen, Song Yaqi. Zhu Yongli. Information platform of smart grid based on cloud computing. Automation of Electric Power Systems. Vol. 34 (2010) No. 22, p. 7-12.

[5] Xindong Wu, Xinquan Zhu, Gongqing Wu, Wei Ding.”Data ming with big data,”IEEE Transactions on Knowledge and Data Engineering, 2014, p.97-107.

[6] Huang Jinxin, Niu Lin, etc. Technical aspects and case study of big data based condition monitoring of power apparatuses. Power and Energy Engineering Conference (APPEEC), Hong Kong, 2014, p. 1-4.

[7] Qi Kaiyuan, Zhao Zhuofeng, Fang Jun, Ma Qiang. Real-Time processing for high speed data stream over large scale data. Chinese Journal of Computers. Vol. 35 (2012) No. 3, p. 477-490.

[8] Meng Xiaofeng, Ci Xiang. Big Data Management: Concept, Techniques and challenges. Journal of Computer Research and Development. Vol. 50 (2013) No. 1, p. 146-169.

[9] Shu Hongchun, Sun Xiangfei, Yu Jilai. A survey on the application of rough set theory in power systems. Automation of Electric Power Systems. Vol. 28 (2004) No. 3, p. 90-95.

[10] Kuang Hailan, Liu Xinhua, Chen Zhong, Jiang Tao. The application study of rough set theory in power system data mining. Journal of Hengyang Normal University. Vol. 27 (2006) No. 3, p. 70-71.

[11] Shu Hongchun, Sun Xiangfei, Si Dajun. A study of fault diagnosis in distribution line based on rough set theory. Proceedings of the CSEE. Vol. 21 (2001) No. 10, p. 73-82.

[12] Dong Bo, Qiu Jie, Zheng Qinghua, et al. A novel approach to improving the efficiency of storing and accessing small files on Hadoop: a case study by PowerPoint files. In Proceedings of IEEE SCC, Miami, 2010, p. 65-72.