

# Research on Application of High - efficiency Data Mining to Electric Power Data under Big Data

Weichun Ge, Zhenjiang Lei, Zhao Li, Xiaolei Hu and Jinyuan Liu  
State Grid Liaoning Electric Power Co., Ltd. shengyang 110006

**Abstract**—With the continuous construction of smart grid, power companies produce, transport, sales and management and management have a mass storage, heterogeneous forms, real-time processing of large amounts of data. How to make full use of and analyze these data, rapid access to valuable information of current electric power big data processing is an urgent need to break through the difficulty. On the analysis of the current electricity big data, on the basis of the status quo of the application research on building the intelligent power big data mining technology efficient analysis, from the power of big data analytics integration management, data analysis, data processing, data show four aspects discussed in accordance with the electric power enterprise development needs large data key technology choice. Big data analysis technology in the wide application of electric power industry will bring the change of the industry, the development of smart grid to a new stage.

**Keywords**-*electric power information; large data; data analysis and mining; smart grid; feature extraction*

Power system as the economic development and human life depend on the energy supply system, also has the typical characteristics of big data [1]. Power system is one of the most complex artificial systems, its geographical location are widely distributed, real-time balance, huge Numbers of transmission energy, generating electricity, electricity transmission speed can reach, highly reliable communication scheduling, real-time operation never stops, the significant failure moment expand etc, these characteristics determine the power system operation produce large number of data, rapid growth, type is rich, full compliance with all typical features of big data.

McKinsey report predicts that, worldwide, big data analysis scheme is widely used to bring \$300 billion in annual electricity cut [2]. Power efficient utilization of big data can be geared to the needs of industry inside and outside provide value-added services to a wide range of high value-added business, for the electric power enterprise profit and control level has a high value. Grid industry expert analysis, said when the data utilization rate increase 10%, increased by 20% ~ 49% of the profits can make the power grid. By analyzing the historical data, the service company development decision and guide the scientific development of practice, monitoring company scientific development process is the most commonly used method of the company. Priorities of liaoning has been attaches great importance to the analysis of decision-making power, constantly perfect analysis and decision system, expanding the area of analysis and decision, analysis and decision method, the analysis of the

decision-making level improved significantly, the reform and development to provide strong support for the company. At present, the company is at an important period of strategic opportunities, management GongJianQi transition and reform, analysis and decision work face the situation is undergoing profound changes. Range is more and more wide application field, on the one hand, analysis and decision analysis and prediction of the user to precision demand is higher and higher, affect the results of analysis and decision factors also more and more complex, the data source is more and more extensive, diverse and data volume increased; On the other hand, the power industry after years of development, especially the acceleration of information construction, accumulated the massive industry data, however, all these data in data warehouse, did not play its real value. Due to the large volume of data, type, update quickly, makes it hard for traditional data mining techniques to deal with it, explore its intrinsic value.

This paper focuses on the analysis power data and efficiently under the technology of data mining application research on key technology framework. Then, this paper expounds the power big data analysis in four core key technology, the data processing technology, characteristic analysis, algorithm technology, display technology. Finally, combining with the part of the grid business application scene shows the value of big data analytics explore related applications.

## I. POWER BIG DATA OVERVIEW

In general, big data refers to can't within tolerable time with traditional information technology and the hardware and software tools to perceive, acquisition, management, data collection, processing and services [3]. According to the characteristics of the electric power industry, electric power big data mainly comes from electric power production, management, operation, smart grid, the three areas.

### (1)Electric power production data

Electric power production is a major source of big data, its information system covers electricity, maintenance, security, three main areas of business. Its data type consists of whole life cycle of real-time production data and equipment of two kinds of data, information system mainly includes the enterprise asset management system, technical supervision, performance analysis, real-time database, boiler life and consumption as well as the management, tour the tally, safety supervision, equipment reliability system, etc. On the analysis of such data is mainly focused on how to use historical information to aid in the production of electricity and

equipment maintenance.

(2)Power management operating data

Electric power enterprises need a lot of production and operation of a business support, need across units, across different professional, business data analysis and mining, and in a variety of visual effect way for user friendly interaction with the data. As a result, the analysis of such data using mainly adopts data analysis tools, in the form of multidimensional.easy to understand the data view, all kinds of business activities provide decision-making information for the enterprise.

(3)The smart grid data

As the electric power industry depth of fusion of industrialization and information, the smart grid will carry the power flow, information flow, business flow, power grids and superposition of electric power information network users will occur, the electric power enterprise's overall value will continue to jump, the electric power industry has in the era of big data congenital advantage of the depth data mining and analysis. The smart grid data sources, mainly is the ubiquitous sensor network through high-speed communication networks into operation dispatching center, electrical power using measurable controllable, make the electric power system is more clean, efficient, safe and reliable.

II. THE TECHNICAL FRAMEWORK OF POWER BIG DATA EFFICIENT ANALYSIS MINING

A. The Power of Big Data Integrated Management

The electric power enterprise data integration management technology is from 2 or merger Multiple application system data, to create a more versatile.The process of enterprise applications. From the point of view of integration, is different Source, format, characteristics and nature of the data on the logical or storage Medium organic concentration, for the system to store a series of subject-oriented, integrated, relatively stable, reflects the historical changes of the data collection, to provide comprehensive data sharing system. Electric power enterprise integrated management technology is to solve the electric power enterprise between the various systems of the internal data redundancy and information island [4].

Power big data integration of data management technology, including relational And the relational database technology, data fusion and integration technology, number According to the extracting technology, filtering technology and data cleaning, etc. [5]. Big data An important feature is diversity, which means that the data source Its broad, data types is extremely complex, this complex data environment Big data processing to bring huge challenges, to deal with large data, first must carry on the extraction and integration of data source of data, extract entities and relationships, after association and aggregation unified structure is used to store the data, the data are required for data integration and extracting cleaning, to ensure data quality and reliability of [6].

B. The Technical Framework of Efficient Analysis Mining

For from the national grid company 4 big platform and external data sources of multi-source heterogeneous data, from the data source layer, characteristic analysis layer, service layer, and validation, etc to power large data intelligent efficient analysis mining technology framework. Throughout the technical framework, power large data respectively through the analysis of characteristics of data, data analysis, data were calculated based on memory parallel analysis technology, from the traditional electric power data analysis of characteristics of the ascent to full data processing, data model, high-performance interactive analysis, the final formation of power big data parallelism analysis service system, and in production overhaul management and forecasting applications verified technical applicability. The overall technical framework is shown in figure 1.

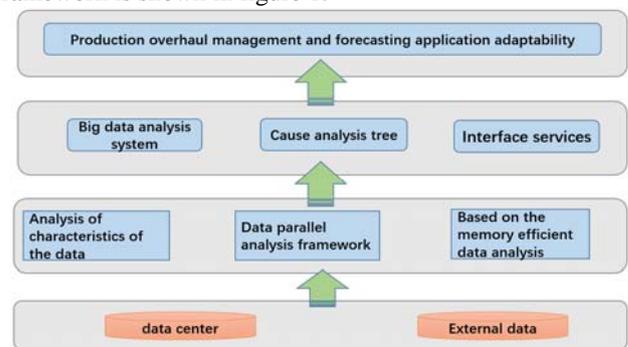


FIGURE I. BIG DATA SYSTEM FOR ENERGY INTERCONNECTION

One data source layer mainly describe the power source of data for large data, mainly includes two parts: 1) the four big national power grid platform of structured, semi-structured, and unstructured data and geographic information data space; 2) weather, social, mobile and other external data.

Characteristics analysis layer describes the data feature extraction method, the method of building data model and data parallel analysis framework. Service layer mainly through service architecture and service interface to build power large data parallel analysis service system.

Application layer overhaul monitoring and forecasting for production application scenarios, by calling the corresponding interface to complete the production of heavy data real time analysis.

III. KEY TECHNOLOGIES IN DATA ANALYSIS

With further data mining and analysis technology, further improve the data processing ability, making it possible to mass data of online monitoring, people face is no longer a historical data but is happening ", the "data production process; The current and future stage, big data analysis is developing towards effective forecast this goal, the analysis of "what is going to happen it will be possible to [7]. On grid business application system in all kinds of static and dynamic data information, for the research of intelligent analysis technology, rapid access to valuable information from the data, will be conducive to solve the grid business application system of the

data in the real-time calculation, on-line analysis, interactive services, concurrent tasks scheduling efficient processing.

*A The Power of Big Data Processing Technology*

Power of big data processing technology including distributed computing technology, memory computing technology, streaming technology, etc., these three technologies applicable objects and solve the main problem is shown in figure 2. Distributed computing technology is to solve large-scale data distributed storage and processing. Memory computing technology is effective to solve data reading and processing of online real-time computing. Stream processing technology is in order to get real-time, the speed and scale of uncontrolled data.

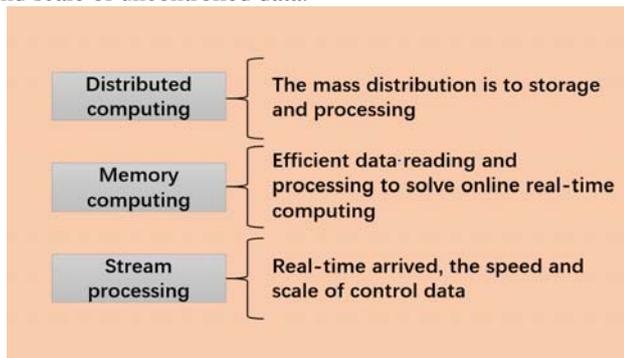


FIGURE II. DATA PROCESSING TECHNIQUES OF ELECTRIC POWER BIG DATA

*B Electricity Power of big Data Characteristics of Big Data Analysis Technology*

*1) Based on the data model of the entity and attribute of the data object*

As abstract extraction, the characteristics of large data of electric power, the need for each type of data modeling in the first place. Accompanied by the emergence of vast amounts of unstructured data need to be able to face big data analysis is a structured, semi-structured, and unstructured data provide a unified service, good data model is the realization of the bottom of the big data unified analysis support. Power big data contain structured, semi-structured, and unstructured data, has obvious characteristics of the business object, through to the business object entity modeling can generate data object, through the records, and records the abstract analysis of the relationship between business objects, can generate the data content of the temporal dimension, spatial dimension, association rules, such as attribute characteristics, is also the power core characteristics of large data, so should the data attribute feature oriented design power big data model. Other behavior characteristics, through the analysis of the data object entity object entity behavior features associated with an increased property characteristics, at the same time, considering the background data and the external factors such as one's profession, finally to classification of all attributes, building put forward object oriented entity, prominent feature attributes, support the general power big complex retrieval data model.

*2) Big data feature extraction based on the data model*

For interdisciplinary, cross types of multi-source heterogeneous electric power data, need to study the data feature extraction method, in power, on the basis of large data model using singularity detection, parameter training, spatial econometric analysis of key technologies, such as at the completion of multi-source heterogeneous data dimension reduction, cleaning after correction and index, using the method of linear and nonlinear characteristics of data extraction, finally to obtain the power of big data attribute feature dimension and association rules. The technical route is shown in figure 3.

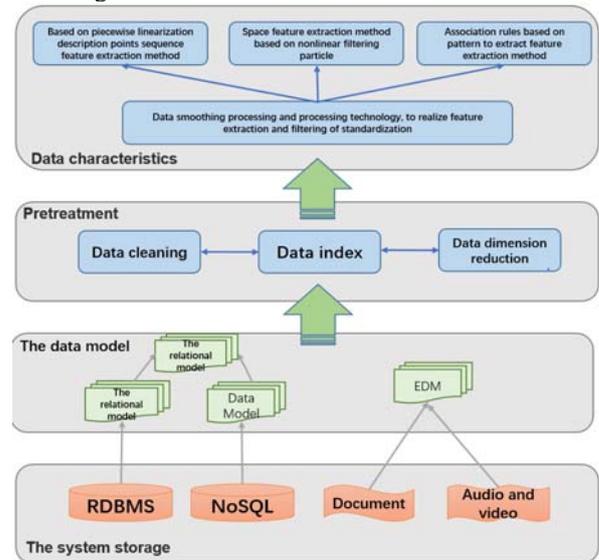


FIGURE III. FRAMEWORK OF FEATURE EXTRACTION FOR POWER BIG DATA

*C Electricity Power of Big Data Algorithm of Large Data Analysis Technology*

Big data technology is the fundamental driving force data, the signal can be converted to the data analysis of information, the information extracting of knowledge, knowledge to contribute to decisions and actions [8]. With the power of big data analysis technology can be from the power system of huge amounts of data to identify potential modal and law, provide decision support for decision makers. McKinsey think that can be used for the key technology of big data analytics is the result of statistics and computer science and other disciplines, including correlation analysis, machine learning, data mining, pattern recognition, neural network, time series forecasting model, a variety of methods, such as genetic algorithm [9].

Big data research study, different from the traditional logical reasoning is a huge amount of data to do a statistical search, classification, comparison, clustering analysis and induction, therefore inherited some characteristics of statistical science, such as statistics the data correlation or relevance of attention, the so-called "correlation" refers to two or more than two variables between the values of some regularity [10]. "Correlation analysis" purpose is to find out hidden in the data

set of networks, general with support, credibility, reflect the correlation parameters, such as interest in degrees.

Big data correlation analysis method, based on large amounts of sample, not using the method of random such shortcuts, and adopt the method of analysis of all data; Big data of simple algorithm is more effective than small data of complex algorithm, the result is faster, more accurate and less susceptible to interference, so he pointed out that based on correlation analysis of prediction is the core of big data. Focus on big data that does not pay attention to causality correlation analysis method, the scientific research way of thinking of a major shift, the late Turing award winner Jim gray data intensive scientific research, "the fourth paradigm" proposed by the big data research three former paradigm (theoretical science, computing science and experimental science) in isolated, alone, as a kind of research paradigm, it is because of the research methods of different from the traditional research on the basis of the mathematical model of [11].

Power big data analysis technology, in essence, belong to the traditional data mining technology in the new development of huge amounts of data mining, but the big huge amounts of data, the characteristics of high-speed growth, diversity, and not only contains structured data, also including semi-structured and non institutional data, so many of the traditional data mining method of processing small data is no longer practical. Big data under the environment of data mining and machine learning algorithms, can from three aspects: 1) from the management of big data and sampling, the Angle of feature selection, the big small digital data; 2) under the big data clustering and classification algorithms, such as least squares support vector machine (SVM) based on conjugate degree (further squares support vector machine, the LS - SVM), random extensible Fuzzy C - Means (FCM), etc. 3) to carry out the big data parallel algorithm, the traditional data mining methods through the parallelization, application of knowledge to the big data mining, such as machine learning and knowledge discovery based on graphs.

#### *D Power of Big Data Power of Big Data Display Technology*

Electric power of big data display technology including visualization technology, space information display technology, such as historical flow display technology [12]. Flow visualization technology, space information technology, history shows technology from three different aspects explain the power big data show the rich connotation of technology. With the power of big data show technology can help managers more intuitive and accurate understanding of the meaning of the power system data expression, to understand the operation state of power system.

Visualization technology is widely applied in the state of the power grid real-time monitoring, improve the automation level of the power system. Future power system visualization can be combined with the relevant theories of complex networks in power grid automatic hierarchical partitioning, the stationing, intensive research, explore law of power grid at a deeper level and contact.

Spatial information display technology is mainly manifested in the grid parameters and the combination of

geographical information system, containing substation 3 d display, such as virtual reality technology. Will power distribution equipment management combining with geographic information system, is advantageous to the power grid management intuitive understanding of the equipment, and for its decision to provide the latest geographic information. Using spatial information display technology in the transformer substation project design can save time, resources, costs and bring huge benefits to electric enterprises.

Historical flow display technology embodies in the management of the historical data of power grids and display. In power system, the application of deep analysis is often based on historical data. For real-time monitoring of manufacturing grid data, planning data and load forecasting data, through the historical flow display technology, can be mapped data and predict the development trend of the future trends; Through the historical stream playback display technology, can simulate and evolution history major event occurs, the knowledge and rule of potential mining historical events.

#### IV. BIG DATA ANALYSIS IN THE APPLICATION OF THE SMART GRID

##### *A. Big Data Analysis in the Application of the Grid Disaster Early Warning*

As the grid increasing complexity and changing the natural environment, chain disastrous accidents in power system, most of these catastrophic accident chain begins with a component fault system. Large-scale blackouts in initial stage is often a small amount of component failure, in expanding the phase is closely linked with the weak link of power system, so from the perspective of the overall prevention, through the technology of data to identify weak links of power network to improve the reliability of power system, reduce the probability of large-scale blackouts have important significance. A typical grid disaster early warning system framework based on big data as shown in figure 4. The framework for grid topology data, geographic information data, the data for the data grid operating, the algorithm model layer library criterion were obtained through the criterion of evaluation indicators, through the analysis of the cause search engines to get cause tree, through the calculation of evaluation indexes for library evaluation indexes. On this basis, in the business layer through the weak area identification, analyzing the causes of weak area, and combined the criterion layer of library evaluation index, comprehensive early warning results are given.

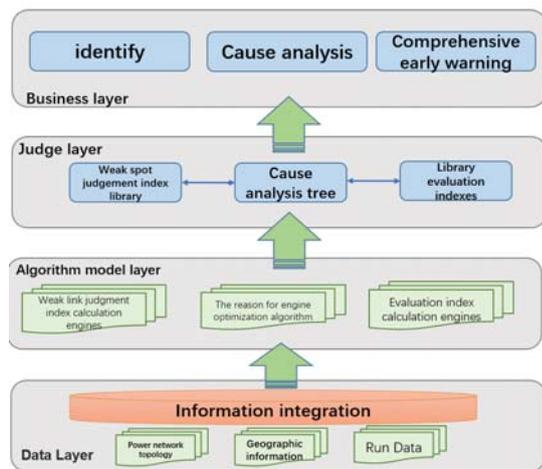


FIGURE IV. DISASTER WARNING SYSTEM ANALYSIS OF POWER SYSTEM BASED ON BIG DATA

**B. Big Data Analysis in User Security Analysis of the Application of Diagnosis**

Important user power supply security big data analysis diagnosis technology is a traditional data mining techniques in the new application and development of huge amounts of data mining, but because of the large amount of data, fast growth, wide variety, and structured, semi-structured, and unstructured data coexist, traditional data expression model and method of shallow model is usually a simple learning, cannot be carried out on the characteristics of the underlying data abstraction, learning efficiency and precision.

Based on big important user data can be classifying user data analysis from three aspects: 1) real-time data, according to the real-time data can be divided into real-time and off-line analysis. 2) the data size, according to the size of the data can be divided into memory, PB and mass level. 3) algorithm complexity of the business, for example, according to two different users voltage quality and reliability analysis, related to the amount of data and algorithm complexity have very big difference, voltage quality analysis of the traditional data analysis methods can be achieved, and the reliability analysis needs comprehensive equipment running state, power supply topology structure, running environment, etc factors, finally it is concluded that the user value reliability. In addition, the most important user power supply safety accident begins with system weak link some equipment components of fault, causing the user a series of fault, eventually lead to the power supply safety accidents. Therefore, in the mining equipment failure events at the same time, also need to analyze the user power supply system of weak links. Based on multi-dimensional state information and data analysis method, the basis of data mining, finally realizes the user electricity safety risk identification.

**C. Big Data Analysis in the Application of Power Grid Production Overhaul**

Production overhaul as a key link power grid operation, involving equipment (facilities) running status, fault defects analysis and problem. Production equipment overhaul project

implementation to the repair order as the main line, through the ERP system supplies, services, the project schedule, funds, such as process control, to achieve a single equipment maintenance cost of the collection. Through data analysis, in order to plan annual, professional categories, professional subdivision, voltage grade, cost of raw materials, such as repair, the total costs for the dimensions. Repeated analysis of the same equipment overhaul, overhaul project cost of raw materials production, repair construction costs in each unit, see the equilibrium degree of professional category, grasp the cost of raw materials and the repair of proportion, and carries on the forecast, the specific analysis process is shown in figure.

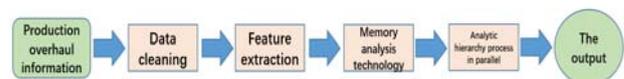


FIGURE V. PRODUCTION OVERHAUL INFORMATION ANALYSIS AND MINING PROCESS

1) in view of the data acquisition system failure or special events may result in data loss, burr and distortion of the problem, first of all, use the big data pretreatment such as empirical correction method in data cleaning, and then for feature extraction.

2) using the production overhaul the historical data of feature extraction can be trained and overhaul cost of raw materials, the repair of the data mining model is established, and based on the analysis of the Spark memory technology to realize the above algorithm, so as to ensure the real-time performance of production overhaul cost prediction.

3) overhaul parallelization of analytic hierarchy process (ahp) is adopted to improve the production management and prediction, and to modify the result.5.

**V. CONCLUSION**

This article focuses on analysis power data and efficiently under the technology of data mining application research on key technology framework, this paper expounds the power big data analysis in the key of the four core technology, the data processing technology, characteristic analysis, algorithm technology, display technology. Big data analysis techniques from big data governance and sampling, large data feature selection, big small digital data, data classification algorithm, parallel data mining,etc, to carry out the research.

With the smart grid, three sets of five construction, business system and data center of rapid accumulation of large become big data, analysis of the construction of the system model, computing speed, ability, and other aspects are obvious effects, but also bring unprecedented rich condition data analysis. Big data analysis technology can create value in five aspects, respectively is: create transparency, let stakeholders easier access to information; Through experiment to find demand, exposed the variable factors and improve performance; According to the customer demand subdivision crowd; Through an automated algorithm to replace or support for decision making; Innovative business model, products and

services. Big data technologies have become more widely used in the field of business and create a huge commercial value, but its application in electric power system has just started, so combining technical advantage of big data and the application requirements of power system, exert the power of big data value, will bring new development opportunities for the smart grid construction. Power enterprises should firmly seize the opportunity, from the data policy, talent training, the key technology research and development level, such as comprehensive to the development of the technology of data analysis.

## REFERENCE

- [1] Gao S, Chau K T, Liu C, et al. Integrated Energy Management of Plug-in Electric Vehicles in Power Grid With Renewables[J]. IEEE Transactions on Vehicular Technology, 2014, 63(7):3019-3027.
- [2] Ekström J, Koivisto M, Mellin I, et al. Assessment of large scale wind power generation with new generation locations without measurement data[J]. Renewable Energy, 2015, 83:362-374.
- [3] Ahern T, Benson R, Casey R, et al. Improvements in Data Quality, Integration and Reliability: New Developments at the IRIS DMC[J]. Advances in Geosciences, 2015, 40:31-35..
- [4] McKinsey& Company.Big data:the next frontier forinnovation,competition,and productivity[M].New York: McKinsey Global Institute,2011:1-28..
- [5] Wu J, Li P, Qian H, et al. Using correlation and multivariate statistical analysis to identify hydrogeochemical processes affecting the major ion chemistry of waters: a case study in Laoheba phosphorite mine in Sichuan, China[J]. Arabian Journal of Geosciences, 2014, 7(10):3973-3982.1
- [6] Hey T. The Fourth Paradigm – Data-Intensive Scientific Discovery[M]// E-Science and Information Management. Springer Berlin Heidelberg, 2012:1.