Development of a research platform for BEPC II accelerator fault diagnosis

Qi Zhang^{1,2} · Jianshe Cao^{1,2} · Yanfeng Sui^{1,2}

Received: 23 March 2020 / Revised: 8 May 2020 / Accepted: 13 May 2020 / Published online: 16 June 2020
© The Author(s) 2020

Abstract

Background BEPC II is an electron–positron collider with a beam energy of 1.89 GeV and luminosity of $1 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$. Being a complex accelerator, fault diagnostics both in an accurate manner and in time are often challenging during its operation.

Purpose The fault diagnosis research platform is capable of locating and resolving faults on time, thus can largely improve the operation of BEPCII by reducing downtime due to machine fault and consequently satisfying the demand for fault diagnosis of a complex system.

Methods Making full use of the existing large-scale hardware foundation, abundant data resources and previous experiences, a fault diagnosis research platform was built by applying data mining, machine learning and other related algorithms to carry out specific fault analysis on the entire system.

Conclusion The fault diagnosis research platform has been developed with implemented functions of data query, statistics and analysis. Partial fault diagnoses of the three subsystems have been realized. The 4W1 power supply fault has been successfully discovered, and effective solutions have been subsequently provided.

Keywords BEPC II · Integration framework · Fault diagnosis · Data mining · Machine learning

Introduction

The Beijing Electron Positron Collider II (BEPC II) is a large scientific project in China, consisting of a 202 m long linac, a transportation line, two 237.5 m long storage rings, the Beijing Spectrometer (BES), the Beijing Synchrotron Radiation Facility (BSRF) and a computer center [1]. Being a complex system, BEPCII consists of considerable amount of data, such as real-time data and historical data, as well as various kinds of software and hardware parameters. However, it lacks dedicated tools for fault data analysis and research.

Since 1980s, fault diagnosis researches on large particle accelerators have been paid much more attention from scholars and experts worldwide. In 2014, the British Spallation

Neutron Source built the First-Line Diagnosis tool, FLDt. With 220 fault paths, it uses a hierarchical fault flowchart to evaluate and feedback through wireless or wired connections and finally present to the user in a web page [2]. The application of FLDt maintains the sustainability of the equipment and facilitates periodic trend analysis, which is easy to understand the physical characteristics of the machine, improves the machine diagnosis rate, reduces fault-finding time and so on. In 2016, Shanghai Institute of Applied Physics used Matlab to read the individual signals of some devices on the beamline, to realize research on the alarm of beamline equipment, but it only used Matlab neural network to conduct the alarm analysis of the data, which is a preliminary attempt of fault diagnosis [3]. Its structure model object is relatively simple, and only suitable for individual running state, but not for the whole system structure with a large amount of data.

The fault diagnosis of massive operation data is mainly divided into three aspects: knowledge driven, data driven and value driven. Based on the investigation of a large number of relevant data, the design and implementation of the research platform for BEPC II accelerator fault diagnosis were completed by using the value-driven fault diagnosis method.

✉ Yanfeng Sui
syf@ihep.ac.cn

¹ Key Laboratory of Particle Acceleration Physics and Technology, Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China

² University of Chinese Academy of Sciences, Beijing 100049, China

Fig. 1 SSM frame structure model

According to BEPC II operation data, the platform completes data acquisition, establishes historical database, develops web application procedure, builds Hadoop cloud platform, adds Kettle, HDFS, Elastic Search and other technologies. It does self-test analysis, classified recording and alarm display of the historical data and cleans the processed historical data to form statistical charts of the data operation state. With the adoption of modular design, it has classification on BEPC II statistical data of subsystem and extraction for modeling, and then, according to the different states, it makes related sample data set. It can realize the fault diagnosis and analysis of 4W1 power supply with the adoption of DBSCAN (density-based spatial clustering of applications with noise) algorithm [4].

System framework design

BEPC II fault diagnosis research platform mainly uses Java; the system framework includes four parts: BEPC II historical database, data acquisition and query system, data statistics system and data analysis system.

BEPC II historical database can organize, store and manage all data, and users can add, query, update, delete and other operations the data in the file. The system can realize the data extraction, storage and retrieval functions, and implement data cleaning and migration, and complete data preprocessing. Through the modular design of this system, it establishes standard interface, separates the logical field, selects the method for analysis and realizes targeted fault diagnosis analysis.

System development

BEPC II fault diagnosis research platform uses the SSM (Spring MVC + Spring + MyBatis) integration framework as the model for development [5]. As shown in Fig. 1, the frame structure model consists of three main frames: Spring MVC, Spring and MyBatis with Spring frame as the core to integrate other frames. The development environment is

Windows system, Java version is JDK 8, the server is Apache Tomcat, the foreground interface is jquery-easyui-1.2.6, and the theme is ui-sunny.

Spring is a lightweight container framework of IoC (inversion of control) and AOP (aspect-oriented programming), which can integrate all major frameworks and reduce software dependence. Spring MVC separates the controller, model object and distributor, which makes them easier to customize. MyBatis is a semiautomated ORM (object relational mapping) framework, which is less difficult to develop than Hibernate, and can be written manually as SQL (Structured Query Language). Therefore, users can freely control the size of fields queried and returned by association and can easily modify them in XML (Extensible Markup Language) files. JQuery EasyUI is a set of jQuery based foreground plug-in assemblies, providing a complete set of components for creating cross-browser web pages which include powerful data grid, tree grid, panel, combo, etc. The page supports all kinds of themes, is lighter and supports two rendering methods—JavaScript method and HTML marking method. And it supports HTML5.

The software functions are mainly divided into presentation layer, business logic layer, data persistence layer and domain module layer. The presentation layer is based on the Spring MVC frame, controlled by Spring MVC development page as view container, and JQuery EasyUI development view page to control business jump. MyBatis is used as the data persistence layer to persist the database data and provide data support; as the core management, Spring manages Spring MVC and MyBatis, providing JQuery EasyUI page injection and data injection. The Spring framework uses Spring IoC to inject instances into interfaces to realize loose coupling.

The whole framework model realizes mapping of data tables and configuration of database connection information through configuration XML files, configuration Java Bean and automatic annotation functions in the application-Context.xml files realize dependency search and instance injection.

Fig. 2 BEPC II historical database table structure

Fig. 3 The structure of data acquisition and query system

BEPC II historical database

This BEPC II historical database is based on MySQL database, adopts table space and table partition technol-

ogy and extends BEPC II existing database table structure. Currently, the data archiving tool used by BEPC II local station is Channel Archiver which is a text file-based data archiving tool in EPICS system. Its main function is to

Fig. 4 Flowchart of data statistics

receive and store the physical variables data transmitted from IOC through Channel Access [6]. On the BEPC II local server, `mysql_schema.txt` is applied to generate Archive database, corresponding `ALARM_MYSQL.sql` and `MYSQL_USER.sql` are applied to generate Archive table structure, and `ArchiveConfigTool` is used to import the written configuration file into MySQL database. It can write the required amount of PV profile and import it into MySQL database.

Figure 2 shows BEPC II historical database table structure. Application Archiver exports and sorts out the operation data from the local station from October 2014 to now (5 rounds in total) with a total size of 3.4T which includes 8 subsystems and 1269 variables.

Data acquisition and query system

The data acquisition and query system use Hadoop for big data analysis. The development environment is Linux SLC 6.7 system; the platform is Hadoop-3.0.3.

As shown in Fig. 3, BEPC II historical data acquisition and query system stores the structured data in different subsystems into the database system after the ETL (extract transform load) process is associated and assembled, stores

Fig. 5 Flowchart of data analysis

the independent documents and so on into HDFS (Hadoop distributed file system), generates a keyword library after word segmentation processing and establishes an index on the basis. After the user sends out the request, the system judges the application type according to the user's demand and calls the corresponding function. Among them, Kettle is used to realize system data extraction, conversion and loading functions, HDFS technology is used to realize distributed storage of archive data. Elastic Search distributed association retrieval technology is adopted to realize retrieval application, and the data mining and analysis functions are realized by Apache Spark [7].

Data statistics system

As shown in Fig. 4, data extraction is implemented by the open-source tool Kettle. Through the front desk interface, physical variables from different data sets can be managed

Fig. 6 Fault diagnosis research platform interface

and data extraction is efficient and stable. The extraction process is implemented through transformation and job scripts, in which transformation completes the data conversion and job completes the control of the whole workflow.

Transformation can realize data migration between the diagnostic platform or databases, export physical variables from BEPC II local databases to files, import large-scale physical variables to historical databases, and clean and integrate physical variables into the diagnostic platform. According to the data filtering conditions, the physical variables data is filtered and converted and then loaded into the storage structure. Besides, job mode or operating system scheduling is used to execute a conversion file or job file, which is deployed on multiple servers in a cluster mode to realize distributed retrieval function, establishing a field mapping table to establish a corresponding relationship between physical variables and extracted data.

Data analysis system

As shown in Fig. 5, the functions of the data analysis system include statistical analysis and some machine learning.

After the requirement proposed by users, it can select the physical variables name and specific time to be analyzed, and according to their relationship, the system can establish logical combination. The system checks the physical variables group and quantity. After confirmation, the system selects the corresponding analysis method for users, sets the corresponding gold parameter and performs the validity check. After confirmation, the system can have a data query request and execute it, and execute the corresponding related analy-

Fig. 7 The sample data set of 20181001-R1OBPM15X-R1OBPM15Y

sis algorithm. After obtaining the analysis result, it can store the model and present diagram report.

Fault diagnosis research platform

As shown in Fig. 6, according to the specific requirements of the system, functions such as account login, modification, management, cancelation and so on, data query (single or multiple physical variables changes with time), data statistics (data set), data analysis (classification, clustering and other machine learning results) and so on are realized. After testing, IE11, Firefox, Google Chrome and 360 browsers can be used for normal access and have good compatibility.

Fig. 8 The sample data set of 20181001-R4OBPM16X-R4OBPM16Y

Data preprocessing

All data are preprocessed first and divided according to variable group time. Since the exported file size cannot exceed the limit of 2G, each group of variables divides the data into different periods such as daily, every 5 days, every half month, every month, every half year, every year and so on according to its actual size and then stores the data. When storing, unifying the data format of continuous analog data is to unify the time and frequency of the data. The unified data recording time is once per second. If there are more than one data in one second in the original data, the average value in one second is taken as the recorded data in this second. If there is only one data recorded for more than one second, the data is recorded according to the next full second of data recording. Because of the limitation of local command characters, the cryogenic (CRYO) data is divided into three groups accord-

Fig. 9 Periodic fluctuations of R1OBPM15 and R4OBPM16 horizontal readings

Fig. 10 R1OBPM15 and R4OBPM16 clustering results

Fig. 11 SRWM:I and SRWT:I clustering results

Fig. 13 SRWM:I and R4OBPM16 clustering results

Fig. 12 SRWM:I and R1OBPM15 clustering results

ing to its variables and each group divides the data in time according to its actual size before exporting.

After the data format is converted and unified, the variable data is modified with different accuracy (basically following the rounding principle) in different data sets according to the variable meaning to ensure the feasibility, timeliness and accuracy of subsequent data analysis. Check the numerical characteristics of the data, such as maximum, minimum, mean, standard deviation and quantile. Find out if there are missing values, and if there are missing values NaN, fill 0 automatically.

First, the physical variables data is read and processed by files, then it is made into sample data set, and its characteristic variables, category variables and so on are determined. The data sets currently processed include R3O:BI:DCCT:CUR, R4O:BI:DCCT:CUR, R1OBPM15, R4OBPM16 horizontal and vertical data, 4W1 main power data and so on.

Figure 7 shows the sample data set at X and Y position of the outer ring BPM15 on October 1, 2018. And Fig. 8 shows the sample data set at X and Y position of the outer ring BPM16 on October 1, 2018. Among them, data points in different colors indicate different density distributions in this data space.

4W1 power supply fault diagnosis

During the last two rounds of BEPC II, R1OBPM15 and R4OBPM16 horizontal readings fluctuate periodically and the fluctuation is about 80% of the normal period. As shown in Fig. 9, the R1OBPM15 horizontal reading is blue and the R4OBPM16 horizontal reading is red. Because of this phenomenon, the specific fault problem is analyzed and solved.

The representative BPM data set from October 2 to 8, 2018, is made, and the DBSCAN algorithm is applied to compare the clustering results with the optimal parameters. The clustering results will divide all the data into different colors, all the points with the same color represent a cluster formed after clustering, and different colors represent different clusters. Then, the color in the clustering result graph is the same meaning and will not be expressed again.

As shown in Fig. 10, with the selection of the optimal parameters $\text{eps} = 0.004$ and $\text{min_sample} = 3$, the left graph

Fig. 14 R1OBPM15 and R4OBPM16 readings in horizontal direction

gets 16 clusters and the running time is 15.48 s; the right graph gets 27 clusters and the running time is 16.44 s.

At the same time, the clustering conditions of other relevant variables in the same period are extracted [8]. As shown in Fig. 11, with the selection of the optimal parameters $\text{eps} = 0.05$ and $\text{min_sample} = 4$, the graph gets 61 clusters, and the running time is 120.60 s. It is found that the main power supply (SRWM:I) and auxiliary power supply (SRWT:I) of 4W1 wiggler magnet have the same type of separated clusters in the same period.

According to this finding, the data set of main power supply (SRWM:I) and R1OBPM15 was made and cluster analysis was carried out to obtain Fig. 12. As shown in Fig. 12, with the selection of the optimal parameters $\text{eps} = 0.1$ and $\text{min_sample} = 3$, the graph gets 20 clusters, and the running time is 12.29 s. The data set of main power supply (SRWM:I) and R4OBPM16 was made, and cluster analysis was carried out to obtain Fig. 13. As shown in Fig. 13, with the selection of the optimal parameters $\text{eps} = 0.1$ and $\text{min_sample} = 3$, the graph gets 27 clusters, and the running time is 16.25 s.

It is found that the large amplitude fluctuation of R1OBPM15 and R4OBPM16 is caused by the large amplitude fluctuation (2%) of 24 h in the main power supply of the wiggler magnet. After the power supply was repaired on October 9, as shown in Fig. 14, the large-scale periodic fluctuation of the orbit in horizontal direction disappeared and the problem was solved. The R1OBPM15 horizontal reading is blue and the R4OBPM16 horizontal reading is red.

Conclusions

A fault diagnosis research platform has been developed for BEPCII. Data mining, machine learning and other related algorithms have been implemented to carry out fault analysis, evaluation of machine performance and research on machine

reliability. These provided a reference direction for fault diagnosis and resolution of the problems. Partial fault diagnoses of the three subsystems have been realized. The 4W1 power supply fault has been successfully discovered, later analyzed and eventually solved.

Funding Funding was provided by Youth Innovation Promotion Association of the Chinese Academy of Sciences (Grant No. 2016011)

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. BEPC and BEPC II, <http://english.ihep.cas.cn/doc/1840.html>, 2017
2. J. Brower, The fault analysis tool for the ISIS Accelerator, https://adweb.desy.de/mpy/WAO2014/Talks/Tuesday/Session_1/3-Brower.ppt, 2017
3. H. Sun, Research of Fault Diagnosis Methods of Beamline in SSRF Based on Neural Networks, Ph.D. thesis, Shanghai Institute of Applied Physics Chinese Academy of Sciences (2016)
4. W.N. Qian, A.Y. Zhou, Analyzing popular clustering algorithms from different viewpoints. *J. Softw.* **13**(8), 1387 (2002)
5. Spring, <https://spring.io/>, 2018
6. EPICS, <https://epics.anl.gov/>, 2017
7. IK Analysis for Elastic Search, <https://github.com/medcl/elasticsearch-analysis-ik>, 2019
8. T. Lin, T.K. Ma, D.Y. Qin, S. Dong, Research on wind turbine fault diagnosis based on improved DBSCAN algorithm. *Mod. Electron. Technol.* **41**(21), 147 (2018)