

DATA MINING APPLIED IN MANAGEMENT OF HEAVY ION ACCELERATOR POWER SUPPLIES *

Huajian Zhang^{†1}, Peng Sun, Qisan Qian, Daqing Gao, Institute of Modern Physics, Chinese Academy of Sciences, Lanzhou, China

¹also at University of Chinese Academy of Sciences, Beijing, China

Abstract

Scientific and effective management of power supplies could reduce the failure rate and improve the efficiency of the heavy ion accelerator. This paper shows how to introduce data mining into the intelligent management of heavy ion accelerator power supplies. A web site platform was developed to collect raw data. The raw data includes many kinds of information about one power supply's life cycle from its development to operation. Among which the failure records are particularly important. According to the attribute that the records are mostly nominal data, R software and SQL Server 2008 Business Intelligence Development Studio were chose as mining tools. R software was used to carry on the statistical characteristic analysis and SQL Server 2008 Business Intelligence Development Studio was used to find out association rules. Useful conclusions have been drawn. This work has laid a solid foundation to further establish the intelligent management system of heavy ion accelerator power supplies.

INTRODUCTION

As special power supplies, the main purpose of accelerator power supplies is to provide specific current for magnets so it can generate the needed magnetic field of the accelerator. The steady running and low failure rate can effectively improve the whole efficiency of the accelerator. The stable operation of the power supplies relies on the continual improvement of the production technology and the scientific management. In recent years, represented by digital power supplies [1], the production technology is quickly developing. While the scientific management of power supplies is in groping stage. So we attempt to apply data mining into the management of heavy ion accelerator power supplies. Data mining is a powerful tool to get useful knowledge and information from huge amounts of data [2]. Researchers at CERN's Large Hadron Collider (LHC) used data mining to analyze 13 million gigabytes and found boson like Higgs [3]. Hefei Light Source (HLS) in China also implemented online analytical processing in their work about accelerator [4].

DATA COLLECTION PLATFORM

Heavy Ion Research Facility in Lanzhou-Cooler Storage Ring (HIRFL-CSR) is the accelerator which could provide highest energy, best quality and a wide range of beams among those built in the Chinese Ninth Five-year Plan period [5]. There are more than 800 power supplies

in HIRFL-CSR. Relying on the experience of heavy ion tumour therapy gained by HIRFIL-CSR, two heavy ion medical machines are under construction in Wuwei and Lanzhou, Gansu province, China (HIMM-WW and HIMM-LZ). And each contains about 160 power supplies. In order to collect information of power supplies in different places, a website platform is designed. So staff of different accelerators can input records via network. There are many kinds of records like parameters, configurations, failure logs and maintaining logs of power supplies.

The sketch of the data collection website platform is showed in Figure 1. Abbreviations PS means power supplies. It follows the B/S structure. Spring MVC architecture is used in the web server application. The architecture is a lightweight Web MVC framework implementing request driving type based on Java. Database management system of excellent performance used in the database server is Oracle 11g.

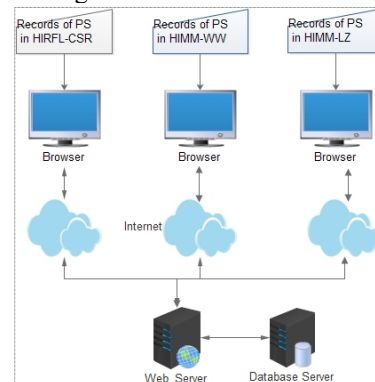


Figure 1: Sketch of the data collection website platform.

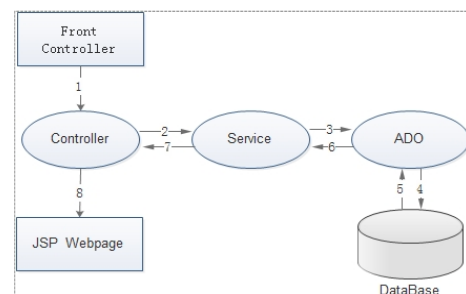


Figure 2: Programming frame of Spring MVC architecture.

Figure 2 shows programming frame of Spring MVC architecture [6]. Step 1, the front controller gets request and data from user URL and calls specific controller on request. Step 2, target controller calls specific server

* email address zhanghj@impcas.ac.cn

method to deal with the data. Step3, target server calls DAO to access database. Step 4, DAO manipulates the database such as selecting, inserting, deleting and updating. Step 5 to step7, result data is transferred to the controller. Step 8, data is organized in the JSP webpage and transferred to the client browser. It's needed to designed different kinds of controller, service and DAO according to service logic so as to realize two-way data transmission between database and client.

DATA MINING PROCESS AND RESULTS

Data mining process framework generally includes data layer, logic layer and presentation layer. Data layer involves raw data selecting and preparing. Data extract-transform-load (ETL), data mining method selection, data mining model establishing and the final realization of data mining compose logic layer. The presentation layer shows data mining result in easy understood way.

In data layer, 1320 power supply failure records from May 8, 2003 to September 6, 2014 in the oracle 11g database were selected and exported to disk file. The main fields of the record are shown in Table 1.

Table 1: Main Fields of Power Supply Failure Record

Field	Data Type	Attribute
PS name	nvarchar(50)	nominal
PS system	nvarchar(50)	nominal
happening time	timestamp	numeric
phenomenon	nvarchar(200)	nominal
handling	nvarchar(200)	nominal
cause	nvarchar(200)	nominal
failure level	nvarchar(10)	nominal
duration	float	numeric

In logic layer, the records were manually processed. We modified the wrong character, standardized expression, combined different expressions with same meaning, and repaired missing data and so on. Then the original data was prepared. For the reason that the data was mainly nominal, we only preliminary analyzed its statistical characteristic and mined its association rules. The goal of the association rule mining was to find out potential relevance among the record fields. The powerful and easy-to-use tools R and SQL Server 2008 Business Intelligence Development Studio were chose.

Data Statistical Characteristic Analysis with R

R is a function language used in statistical calculation and mapping [7]. Open source software R provides abundant statistics and mapping tools. Characteristic analysis of the power supply failure records followed next steps.

Firstly, data extract. The data was processed according to R's restrict. For instance, replace character '#' with other character, convert timestamp into nominal character.

Secondly, the data was loaded to R's dataset.

Thirdly, the command summary() was used to analyse the data. If the field is nominal, R will list the ones which have higher frequency. If the field is numeric, data distribution characteristics will be display. Figure 3 shows the

result. It's clearly that the mean of duration is bigger than the third quartile and the max is very big. That's because there are many '0.00' in the duration field and there are three outliers 3600.00, 4000.00 and 4320.00. Replaced the three outliers with mean 28.53 and analysed again, the mean changed to 19.56 minutes.

PS name	PS system	phenomenon
BL1-BSW : 36	CSRm :271	No current output:207
PS20D : 35	BeamLine2:243	Frequent trip :120
BL2-BSW : 22	SSC :216	One trip :105
PS40D : 18	CSRe :92	Unable to switch:101
SSC-PC : 17	SFC :89	Oscillation :40
PS22Q05 : 16		
duration(minutes)	cause	
Min :0.00	Power System error	:283
1st Qu :0.00	Component aging	:273
Median :0.00	Remote control error	:222
Mean :28.53	Unknown	:138
3rdQu :15.00	Cooling system error	:66
Max :4320	Power and Cooling error:61	

Figure 3: Analysis Result with R

Mining Association Rules with SQL Server 2008 Business Intelligence Development Studio

SQL Server 2008 Business Intelligence is a new development and management tool for business intelligence solution developers. It's a data mining and business intelligence platform. Association rules mining followed below steps.

Firstly, data extract. The original data was further pre-processed to fulfil market basket analysis. The records were split into two tables. Table1 had a field fault number. Table2 had two fields fault number and fault detail. The power supply name, power system, phenomenon, handling, cause, failure level were all fault detail. If fault phenomenon or other detail had more than one element, it would be split into multi records. After data extract, there were 1320 records in table1 and 7210 records in table2.

Secondly, the two tables were loaded into SQL Server 2008. They become tables in relational database. They were related by foreign key fault number.

Thirdly, association rules mining in Business Intelligence Development Studio..

1. Created an analysis services project.
2. Created data source and data source view. Selected the two tables in SQL Server 2008 and created a one-to-many connection between them via field fault number.
3. Create mining model and chose Microsoft association rules. Fault number was set to be key and fault detail was set to be key, input and divivable. Set the parameters of mining method like Table2.

Table 2: Parameter Setting of Association Rules Mining Method

Parameter	Value
MINIMUM_SUPPORT	0.01
MINIMUM_PROBABILITY	0.4
MINIMUM_IMPORTANCE	0

Association rule $A \rightarrow B$, support (A,B) is defined as the proportion of transactions which contains the both items A and B. Parameter minimum_support is the minimum

threshold of support. It was set to be 0.01, so only occurrence number is greater than 1320×0.01 would be considered. Probability, also confidence, with respect to a set of transactions T, is the proportion of the transactions that contains A which also contains B. Parameter minimum_probability was set to the default value 0.4. Importance, also interestingness, is the logarithm of proportion which is the probability quotient of B with A or without A. If importance is 0, A and B are independent. If it is negative number, A and B are negative correlation. So parameter minimum_importance was set to be 0.

4. Deployed and run the mining model, viewed the mining result. Figure 4 shows the dependency network. In the network, if a node has more connecting lines it has greater impact in the whole sample. The network reveals a lot of useful information about power supply failure which can be analyzed to form a report manual by technical personnel. Take the node component aging for instance, when a power supply is unable to switch, the cause is probably component aging. Most failures of power supplies in Beam Line1 and SFC are caused by component aging. Component aging of power supply BL1-BSW in SFC is serious.

[2] Jiawei Han, Micheline Kamber, Jian Pei, *Data Mining Concepts and Techniques Third Edition*, 2012.

[3] Nick Statt, *Higgs: Bringing the power of data-mining to astrophysics, biology*, 2012, <http://phys.org/news/2012-09-higgs-power-data-mining-astrophysics-biology.html>

[4] XIE Dong, LI Wei-min, HE Duo-hui, Application of On2Line Analytical Processing Technique in Accelerator, Atomic Energy Science and Technology, Vol.39, No.5, China, Sep. 2005, pp.451-454.

[5] GAO Da-qing, WU Rong, ZHOU Zhong-zu, CHEN You-xin, *Research and Design of HIRFL-CSR Pulsed Switching Power Supply*, Power Electronics, Vol 37, No.2, China, April 2003, pp.14-16.

[6] Spring MVC programming framework, http://wenku.baidu.com/link?url=UHSR_YeJI67d9ughCk9k4omko3YzV5quNctwiSkFc9-zO-TAy11Te9jFQERhAbHkNwm5_JU8YSII6Uph9bl6zHz-VZqGnuDKnVWwxOOOnne.

[7] Luis Torgo. *Data Mining with R Learning with Case Studies*, 2012.

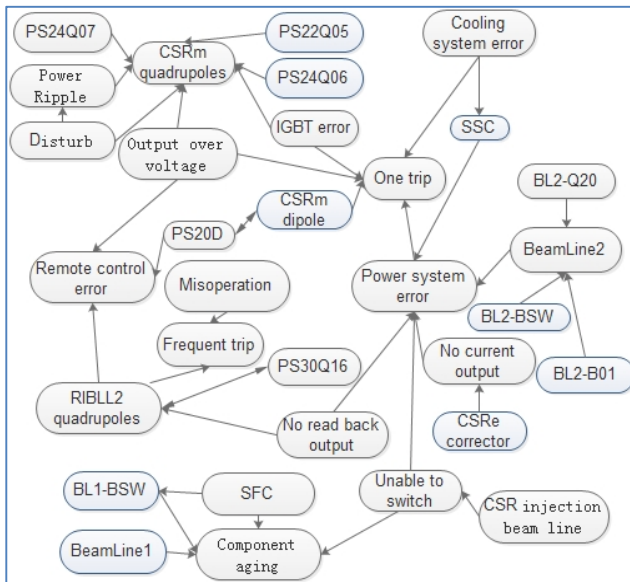


Figure 4: Dependency Network.

CONCLUSION

To introduce the data mining technology would open up a new way of thinking for the scientific intelligent management of power supplies. The potential information can help power supply technical personnel find out feature of power supply's failure. It provides powerful data support to the maintenance, reconstruction of accelerator power supply.

REFERENCES

[1] Rongkun Wang, *Research and Design of Dedicated Digital Power Supply Regulator for Medical Heavy Ion Accelerator*, Doctoral Dissertation, 2013.