Application of Association Rules Data Mining in the Determination the Operation Target Values in the Thermal Power Plant

Wan Xiang, Hu Niansu, Huo Min
School of Power and Mechanical Engineering
Wuhan University
Wuhan, China
e-mail: wanxiang90@sina.com

***Abstract*—As the parameters from the thermal power plant DCS (Distributed Control System) data have closely relationships, the association rules data mining is tried to be applied in the determination the optimization values in the thermal power plant. A model is built by the mining tools Clementine 12.0 based on the GRI (generalized rule induction) algorithm. The analysis on the model shows that the values of controllable operating parameters can be determined in order to improve the thermal efficiency by data mining. Thus such method is helpful for the actual operation in the plant.**

Keywords-association rules, data mining, the operation target values

I. INTRODUCTION

The operation optimization of thermal power units is an important means to improve the thermal efficiency but it is a complicated systematic project [1]. By definition, the operation target values reflect optimum operating parameters and performance indicators the unit can achieve in the current operating conditions, and it provides theoretical guidance for the economical operation and equipment innovation for energy conservation of the thermal power plant [2].

Therefore, it has a strong practical significance to determine the operation target values reasonably.

At present, there are there 3 ways to determine the operation target values: design value way, optimal operation test way and thermodynamic calculation way [3]. The value calculated by the first way is the values that manufacturer calculated by empirical formula and fitting models under the design conditions. The values calculated by the second way are calculated by correction and fitting models under the conditions from the thermal performance test. But the values calculated by the first two ways are fixed values, so they are inaccurate when working conditions have changes. The values calculated by the third way affect by the model greatly, and they are the ideal values which could not reach in the actual operation [4].

On the other hand, many thermal power plants have used DCS (Distributed Control System) with the rapid technology development. These systems can store large amounts of operation data. As traditional data analysis means is just a simple classification and statistics of these data, inability to reveal the rules hidden behind the data, resulting in a waste of resources of data. So a powerful data analysis tools to analyze DCS data is necessary [5].

As an important branch of data mining, association rules data mining describes the correlation between the objects in the database [6]. It is very significant to guide the unit optimized operation that a method which can determine the operation target values can be found in DCS data using data mining.

II. THE BASIC PRINCIPLE OF ASSOCIATION RULES DATA MINING AND BUSINESS UNDERSTANDING

A. *The basic principle of association rules data mining*

For thermal power plants, all levels of the system are closely linked. A parameter changes will spread to other parameters. And there is a complex relationship between operating parameters, load conditions and performance indicators. The relationship can be described as an association rule generated by correlation analysis.

Typically association rule has the following form: A→B，(s, c)

A and B are disjoint subsets of the same data set, s and c are the support and confidence of the rules. They separately describe the usefulness and certainty of the rule. Exactly that data set has s proportion of the data items contained AYB （both A and B）, and c proportion of the data items have to meet the conditions which “If it contains A that it contains B”.

If an association rule can satisfy the minimum support and minimum confidence, so that the association rule is a strong association rule. For example, a strong association rule is operating parameter: A→ performance indicator: B. If B reflects high thermal economy, A will be made sure is the operation target value.

B. *The business understanding of the study to determine the operation target values*

If the operating parameters correspond to a high value of thermal economic under one load condition can be identified from the thermal power plant DCS data, it can be put as the optimal value under the load condition to guide operation. And the task of this paper is to identify the strong association rules meet this criteria from DCS data. This is also the business understanding of data mining.

III. DATA MINING TOOLS—CLEMENTINE 12.0

A. Use of Clementine 12.0

Clementine is one of three famous Data mining tools in the world. It can help discover and predict interesting and valuable relationships present in the data of users.

B. Interface of Clementine 12.0

Clementine 12.0 provides users with a powerful data mining workbench which are easy to use. It contains 7 node areas and they are sources node area, record process node area, filed process node area, graphs node area, modeling node area, output node area and export node area. The modeling process is connecting nodes of each node area together in a manner.

C. Model algorithm of Clementine 12.0

Modeling is the core of data mining process. There are several algorithms in the modeling node area. Each algorithm has certain intensity for particular types of problems.

IV. DATA PREPROCESSING

A. Data understanding

Data understanding is a preliminary analysis to the data; it can be seen as initial data mining process.

In this paper, the selected DCS data are from the fourth 1000MW unit in Jinghai Power Plant. Data extraction ranges from 0:00 on January 9th, 2013 to 0:00 on January 31st, 2013, and the sampling period are 30 seconds. There are a total of 63,339 pieces of data removing a very small amount of missing values. It's available use Excel node in the source area to import data in the Clementine 12.0. In the node, values and whether missing case of each parameter can be observed, the parameters the data mining required will targeted and the others will be filtered out. Data storage format of the selected parameters are shown in Table I.

TABLE I. DATA STORAGE FORMAT OF THE SELECTED PARAMETERS

Parameters	Unit
Power	kW
Main steam pressure	MPa
Main steam temperature	℃
Heat arc reheat steam pressure	MPa
Heat arc reheat steam temperature	℃
Overheating desuperheating water flow	t/h
Reheater desuperheating water flow	t/h
Vacuum	kPa
Feed water temperature	℃
Terminal temperature difference of No.1 high temperature heater	℃
Terminal temperature difference of No.2 high temperature heater	℃
Heat consumption rate	kJ/kW • h

In Table I, heat consumption rate is performance indicators, calculated by the parameters given by DCS database. Except for the heat consumption rate, the remaining parameters are controllable in the actual operation of the power plant unit. It is a strong guiding significance to the actual operation, so they are selected.

B. Data preparation

Sampling process which is one of the 4 steps of the data preparation should be primarily focused on since these data are taken in the same database and are unified units. Sample node of the record process node area in Clementine 12.0 can achieve this functionality. There are 3 ways to achieve this function, and 1/n way is chosen. 1/n way means choose one piece of data from n pieces of data. As sampling period are 30 seconds, is not too short. It will affect the continuity of the data if n is set too large, so set n as 1.

C. Data cleaning

The purpose of data cleaning is make the data clean and tidy and the main contents of data cleaning include the removal of the handle missing values and noise points [7]. Choose node of the record process node area in Clementine 12.0 can delete some data do not meet the conditions targeted.

D. Variable transformation

Variable discretization is a very important step of variable transformation, and an appropriate discretization can improve the accuracy of the classification of the algorithm in some cases. This paper chooses the time when this unit power respectively is 1000MW, 900MW, 800MW, 700MW and 600MW these five typical loads due to the special nature of plant operation to analysis the issue of determine the operation target values.

For example with 1000MW load, in order to improve the accuracy and relevance of mining, data which load is about 1000MW should all be selected. Here, the data that load ranges from 997MW to 1003MW are selected for data mining analysis.

There are a number of means for variable discretization by the binning node of field process node area in Clementine 12.0. Tiles (equal count) are chosen for variable discretization, and data are divided into 10 intervals.

V. MODELING

A. Selection of the algorithm

Algorithm of association rules mainly are Apriori algorithm and GRI (generalized rule induction) algorithm. Since GRI algorithm extracts the maximum amount of information rules based on an index, can handle both digital input and symbol input field, help find the hidden laws of the data [8], and use the depth-first search strategy algorithm [9]. It's more conducive to the data mining among a large number of data of thermal power plants and is chosen to mine.

B. Modeling process

Access GRI node of modeling node area, choose heat consumption rate as consequent, the other parameters as antecedent, the modeling is begin.

VI. Analysis of Modeling Results

A. Analysis of modeling results when load is 1000MW

The modeling results are shown as Table II.

TABLE II. MODELING RESULTS WHEN LOAD IS 1000MW

Parameters	Number of interval	Interval	Average of interval	Optimization
Power（kW）	7	[1000.30,1000.89]	1000.60	1002.54
Main steam pressure（MPa）	4	[24.93.24.99]	24.96	25.17
Main steam temperature（℃）	7	[597.0,597.2]	597.1	596.2
Heat arc reheat steam pressure（MPa）	10	[4.38,4.40]	4.39	4.39
Heat arc reheat steam temperature（℃）	10	[600.5,602.9]	601.7	601.3
Overheating desuperheating water flow（t/h）	3	[21.39,35.64]	28.52	104.06
Reheater desuperheating water flow（t/h）	10	[46.64,56.77]	51.71	47.42
Vacuum （kPa）	10	[4.79,4.94]	4.87	4.88
Feed water temperature（℃）	7	[297.1,297.2]	297.2	297.1
Terminal temperature difference of No.1 high temperature heater（℃）	4	[-1.6,-1.5]	-1.6	-2.17
Terminal temperature difference of No.2 high temperature heater（℃）	3	[0.9,1.0]	1.0	1.46
Heat consumption rate（kJ/kW • h）	1	[6715.33,6865.92]	6790.63	6715.33

The data from the first two columns of this table represent a strong association rule; the data in the third column are the corresponding interval range. The average of the interval in the fourth column is the values calculated by average method between the upper and lower limits of the interval, and it is more intuitive than the interval. And they can be called the values calculated by data mining. The data of the last column is the set of data corresponding to the minimum of the heat consumption rate. It means they are the values of operating parameters when the heat consumption rate is the minimum. And they can be called the values calculated by the optimal method. By comparison， the heat consumption rate calculated by data mining is little bigger than the minimum of the heat consumption rate, but it has enough support and confidence to support, and the probability to achieve it in actual operation is larger than the minimum of the heat consumption rate. Besides, occasional randomness of the minimum of the heat consumption rate is not excluded. Moreover, the minimum of the heat consumption rate is in the interval of the heat consumption rate calculated by data mining.

B. Modeling results under other loads

According to the above method, changes the power range, modeling results under other loads can be calculated, and compare to the corresponding value calculated by the optimal method respectively, which is shown in Table III.

TABLE III. MODELING RESULTS UNDER OTHER LOADS

Parameters	Mining	Optima	Mining	Optima	Mining	Optima	Mining	Optima
Power（kW）	901.53	898.78	800.15	797.73	700.27	701.73	602.79	602.22
Main steam pressure（MPa）	23.35	23.71	21.73	21.85	19.39	19.11	17.10	16.91
Main steam temperature（℃）	596.9	597.3	596.9	593.9	598.6	594.1	598.9	596.3
Heat arc reheat steam pressure（MPa）	3.86	3.84	3.40	3.40	2.99	3.00	2.54	2.62
Heat arc reheat steam temperature（℃）	599.2	596.0	591.2	589.4	596.8	595.3	586.7	595.0
Overheating desuperheating water flow（t/h）	15.61	13.05	12.48	11.36	10.85	62.70	16.30	17.30
Reheater desuperheating water flow（t/h）	44.19	32.71	18.57	3.10	21.68	1.80	36.07	15.43
Vacuum （kPa）	3.85	4.01	3.80	3.63	3.62	3.97	3.41	3.45
Feed water temperature（℃）	290.1	290.5	283.3	284.22	276.88	276.23	268.1	267.14
Terminal temperature difference of No.1 high temperature heater（℃）	-1.4	-1.9	-2.2	-3.6	-3.9	-4.1	-5.1	-2.4
Terminal temperature difference of No.2 high temperature heater（℃）	0.8	0.3	-0.4	-0.8	-0.3	-0.6	-2.3	0
Heat consumption rate（kJ/kW • h）	6984.36	6870.35	6974.63	6803.64	7168.77	6802.14	7195.75	7024.24

In Table III，the data in the column which header is Mining are the values calculated by data mining, while the data in the column which header is Optima are the values calculated by the optimal method. By comparing the results in Table III and Table II, modeling results when load is 1000MW and modeling results under other loads are similar. The heat consumption rate calculated by data mining is little bigger than the minimum of the heat consumption rate, but the gap is little. And the minimum of the heat consumption rate is in the interval of the heat consumption rate calculated by data mining in most cases. Besides, the values calculated by the optimal method occasionally appears just once or twice in a number of data, the support and confidence is much smaller than the values calculated by data mining. So the values calculated by data mining are more representative, better suited to guide the actual operation. Thus they can be considered as the operation target values.

VII. Conclusion

The association rules data mining is applied in the research in the determination the optimization values for a thermal power plant. When the load of the unit is 1000MW, 900MW, 800MW, 700MW and 600MW respectively, a model is built by the mining tools Clementine 12.0 based on the GRI algorithm. In the model the heat consumption rate is consequent and the controllable operating parameters are antecedent. The analysis on the model shows that strong association rules which both support and confidence are ideal can be found; the data mining provides a practical approach to determine the operation target values of the thermal power plant, and it can effectively control the heat consumption rate in a interval which the average of the interval is small. And the values of the controllable parameters calculated by data mining can be regarded as the operation target values to guide the plant operation directly.

References

[1] Yang Yongping，Yang Kun. Theory and application of thermal power energy-saving potential in the diagnosis [J].Proceedings of the CSEE, 1998, 18(2): 131-134.

[2] Ren Haoren, Li Wei, Sheng Deren.The analysis of target value for thermal power units running under variable conditions [J]. Proceedings of the CSEE, 1999, 19(9): 50-52.

[3] Li Jianqiang, Liu Jizhen, Zhuang Luanying. Application of plant operation optimization based on data mining [J].Proceedings of the CSEE, 2006, 26(20): 118-123.

[4] Zheng Xixi. Target optimization studies of the thermal power plants based on association rules [D].North China Electric Power University, 2011.

[5] Qiu Guoxiang. Studies of target optimization determined for the thermal power plants based on association rules [D].Dalian Maritime University, 2006.

[6] Taniar D, Rahayu W, Lee V, et al. Exception rules in association rule mining [J]. Applied Mathematics and Computation, 2008, 205(2): 735-750.

[7] Hu Keyun, Tian Fengzhan, Huang Houkuan. Theory and application of data mining [M]. Tsinghua University Press, 2008

[8] Mao Xiaojian, Wang Heyong. Research of the application for the generalized rule induction algorithm in analysis of pathogen in children with pneumonia [J].Chinese General Practice, 2013, 16(9): 997-1000.

[9] Xue Wei, Chen Huange. Methods and applications of data mining with Clementine [M]. Electronic Industry Press, 2010.