Application of Data Mining Technology in Intelligent Environmental Protection Systems

Guikang Gao*, Mingbo Xiao and Zhen Wang
[1]Hangzhou Dianzi University, China
*Corresponding author

***Abstract*—Among the pollutants in the power industry emissions, sulfur dioxide so far has the most serious impact on the environment. In the desulfurization process, it is common to have data with complex correlations, high real-time, and huge amount. Data mining has become an important technique to deal with these data, and to facilitate better environmental protection and pollution control over the total emissions. The theme of this paper is to effectively analyze the correlation between data, via data mining technology and data association rules, to improve overall decision-making capabilities, and then to provide a reliable basis for intelligent environmental protection. In this paper, we propose parameters predictive model. We verify the proposed algorithms using actual monitoring data of desulfurization processes, and demonstrate that the application of the models has achieved good performance in desulfurization monitoring.**

Keywords-intelligent environmental protection; data mining techniques; association rules; flue gas desulfurization monitor

I. INTRODUCTION

Nowadays, the technology of networking and intelligent platform remains in the developing stage in China[1]. However, the applications and relevant techniques of IoT in "Intelligent Environment Protection" have made breakthrough in many fields[2], such as the advanced environment surveillance system applied in Shanxi Province and Wuxi City[3], the project of "Intelligent Xiangtan" in Hunan Province[4], the water-quality inspection and early warning system in the project of "South-to-North Water Transfer" and so on.

There are still some disadvantages in the available techniques[5]: First of all, the platform with self-inspection is not capable enough to acquire accurate and comprehensive information. Then, China's current environmental information work has just started and imperfection of the construction in some department leads to an insufficient information sharing and utilization. The problems above indicate that the industry of environment protection in China is still in the early stage of development, and the challenges and opportunities coexist in the application of IoT in "Intelligent Environmental Protection" [6].

The main purpose of the research can be summarized as two aspects. One is to improve the capability and skill of information resources sharing. The other is to enhance the environment surveillance and emergency precaution. "Intelligent Environment" aims to make the work of environmental protection with more automation, standardization and normalization, avoiding "Information Silo", which improves the ability of comprehensive policy-making and service in environmental protection.

II. DATA MINING TECHNIQUES

A. *Step of Data Mining*

Under normal circumstances, data mining technology need to go through problem definition, data collection, data preprocessing, data mining algorithms and other stages[7]. Detailed process is as follows.

In practice, the mining process should be divided into the following two procedures. The first step is to find out in advance transaction data among all frequent item sets of data. Following the first step is generating frequent items to derive strong association rules.

B. *Association Rules Algorithm*

Suppose the data items set $X \subset I$, B and A are numbers of transactions X and Y included in the transaction set D, then the definition of support data sets X is as follows:

$$\text{support}\ (X) = \frac{B}{A} \tag{1}$$

Data set containing both X and Y with all matters is defined by the support of association rules as $X \Rightarrow Y$, denoted[8]:

$$\begin{aligned}\text{support}(X \Rightarrow Y) &= \text{support}(X \cup Y) \\ &= P(X \cup Y)\end{aligned} \tag{2}$$

The number of transactions containing both X and Y is divided by the number of transactions containing X, denoted as:

$$\begin{aligned}confidence\ (X \Rightarrow Y) &= \frac{\text{support}\ (X \cup Y)}{\text{support}\ (X)} \\ &= P(Y|X)\end{aligned} \tag{3}$$

III. Application of Data Mining Technology in Intelligent Environmental Protection System

A. *Necessity of Using Data Mining in Power Plant Desulphurization Monitoring*

Currently, electrical equipment is one of the key factors of the economic development, and it is also energy-intensive industries, so high pollution obviously existed in electric power industry, such as sulfur dioxide, nitrogen oxides, soot and so on [9]. Research shows that many of the pollutant emissions actually need to do a detailed study, and the emissions of SO2 is having serious impact on the environment. The power industry can use smart devices for the integrated treatment of emissions, and converted waste to renewable material. This can not only protect the environment, reduce waste, but also save energy and reduce pollution.

The purpose to analyze desulfurization monitoring is to find in many particularities their common characteristics. Then we can take more targeted initiatives to prevent similar situations. Since the power plant desulfurization monitoring data are characterized by large volumes of data, a wide range of complex relationships between data, high real-time, you need to use data mining method to extract the latent, hidden meaningful and valuable information in desulfurization monitoring data. And then we can extract unknown knowledge in desulfurization monitoring data.

In this paper, we introduce a method of association rules for mining desulfurization monitoring data. The specific process is described as follows.

B. *Select Data and Preprocessing*

The selected data sets in this paper come from a Shanghai Environmental Science and Technology Ltd, aimed at doing experiments and monitoring power plant desulfurization facilities. It sets up four monitoring points in the power plant desulphurization facilities, respectively named WL_DC_01, WL_DC_02, WL_DC_03 and WL_DC_04.The capacity of the power plant desulfurization facilities is 600MW. The power plant desulfurization facilities use limestone gypsum wet to deal with desulphurization, and the time interval is one hour.

It can be seen from the specific conditions and further screened out the main test parameter data. Then take FGD exit sulfur dioxide concentration and the efficiency of desulfurization as the set of filtering and classified attribute. Absorber PH value, and other fourteen properties are used as the main object of study. For convenience, these fourteen properties are set A1/A2/A3/A4/A5/A6/A7/A8/A9/A10/A11/A12/A13/A14, as detailed in TableI, which identifies fourteen attributes as the main object of study, clean out the other attributes parameters.

All the data will be compressed and experience transformation process, and eliminate errors of judgment between different parameters. We can take the extreme value standardization as a demonstration, using the maximum standardized way to make the original data which were normalized to the numerical interval within [0, 1]. Referring to the specific process of standardization: $X_i' = [X_i' - \min(X_i')]/[\max(X_i') - \min(X_i')]$ where X_i' is property monitoring value, $\min(X_i')$ is the minimum of the measurement data, $\max(X_i')$ is the maximum value of the measured data.

TableI Monitoring Parameter of Desulfurization

Monitoring parameters	variable
Unit Load	A1
FGD flue gas outlet concentrations of sulfur dioxide	A2
FGD bypass damper outlet pressure	A3
FGD Flue gas inlet oxygen concentration	A4
FGD flue gas outlet temperature	A5
Desulfurization tower at the outlet of the flue gas volume	A6
FGD flue gas outlet oxygen concentration	A7
FGD booster fan front temperature	A8
FGD bypass damper inlet pressure	A9
FGD flue gas inlet concentrations of sulfur dioxide	A10
Electric current of booster fan	A11
The amount of coal combustion	A12
PH value of the absorption tower	A13
Slurry circulating pump current	A14

C. *Establishment of Data Structure Model of Desulphurization Facilities*

1) Establish Parameters Predictive Model

If classified according to the desulfurization efficiency, the attribute can be selected by association rules, in which evaluating algorithm is CFS, while the search method is Best First. Seven properties can be obtained with the desulfurization efficiency-related parameters, namely, A11,A8,A10,A13,A1,A7,A5. Attribute selection is mainly to reduce the less relevant parameter properties in data set, in order to improve the effectiveness of data mining to generate faster and better structured knowledge.

With the seven parameters related to desulfurization efficiency, we use multiple linear regression equation to calculate and predict the desulfurization efficiency of desulphurization facilities.

$$
\begin{aligned}
Z = 112.1825 &- 0.0068 * A1 - 0.3489 * A5 \\
&+ 0.9752 * A7 - 0.0155 * A8 \\
&- 0.5378 * A10 + 0.0502 * A11 \\
&+ 0.0059 * A13
\end{aligned}
$$

where, Z is the predictive desulfurization efficiency value. Formulating this as the predictive model of desulfurization monitoring parameters data, we can use predictive models to generate predictive value of desulfurization efficiency, and to compare with the actual value of desulfurization efficiency. Determine whether the results are in the permissible error range, and if yes, it means normal operation of desulfurization facilities, otherwise, there may be a fault in the desulphurization facilities. In that case, we need to check in time to avoid excessive emission of pollutants.

D. *Prediction and Usability Evaluation Result*

By comparing predictive results obtained from the WL_DC_01 through parameter predictive model and the actual data obtained, we have the result as shown in Figure I.

In Figure I, the green dots represent the real value of desulfurization efficiency, and the red dots represent the model predictive values of desulfurization efficiency obtained by the parameter predictive model. As can be seen, in the normal operation of desulfurization facilities, the desulfurization efficiency is greater than 95.5%, and the maximum difference between the real values and predictive values is 4.0%, which indicates that parameter predictive model is effective.

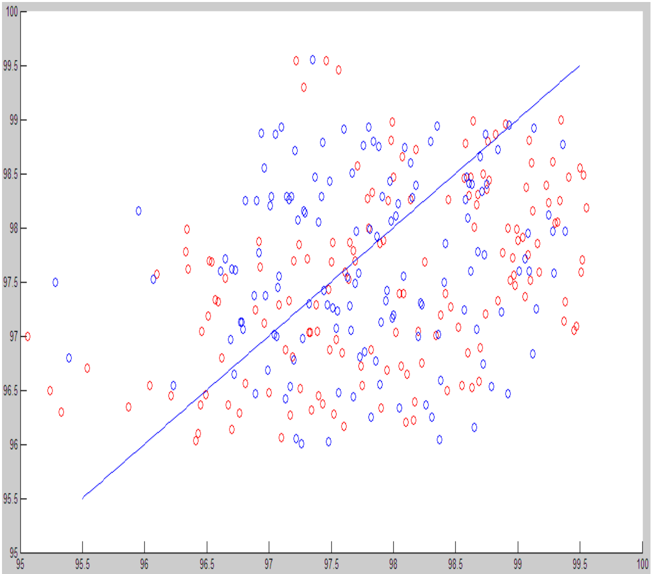

FIGURE I. WL_DC_01 COMPARING PREDICTIVE RESULTS WITH ACTUALLY DETECTED DATA

The comparisons between the predictive value by inputting the desulfurization monitoring data from WL_DC_02, WL_DC_03, WL_DC_04 into the model and the actual data are shown in Figure Ⅱ, Figure Ⅲ and Figure IV.

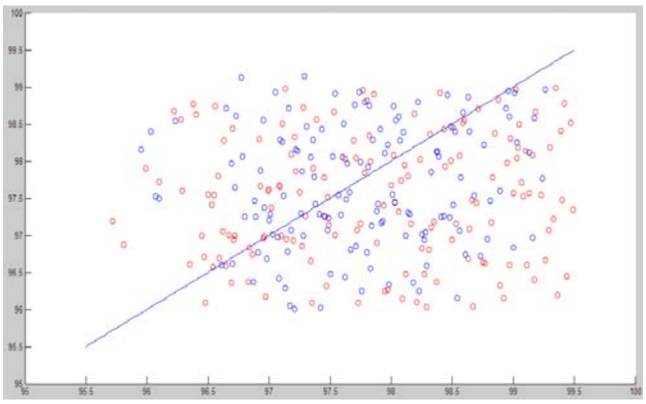

FIGURE II. WL_DHDC_02 COMPARING PREDICTIVE RESULTS WITH ACTUALLY DETECTED DATA

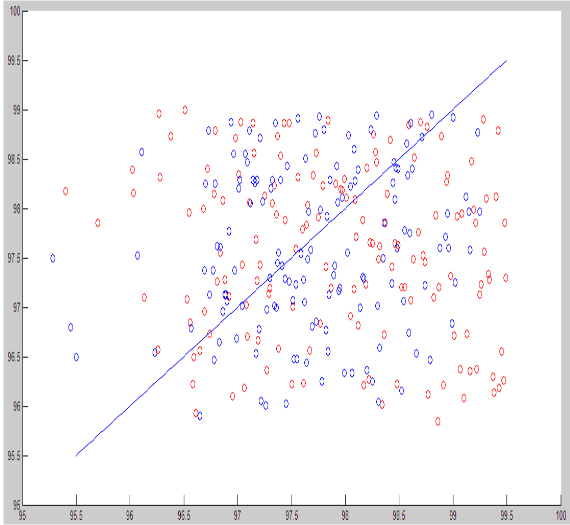

FIGURE III. WL_DC_03 COMPARING PREDICTIVE RESULTS WITH ACTUALLY DETECTED DATA

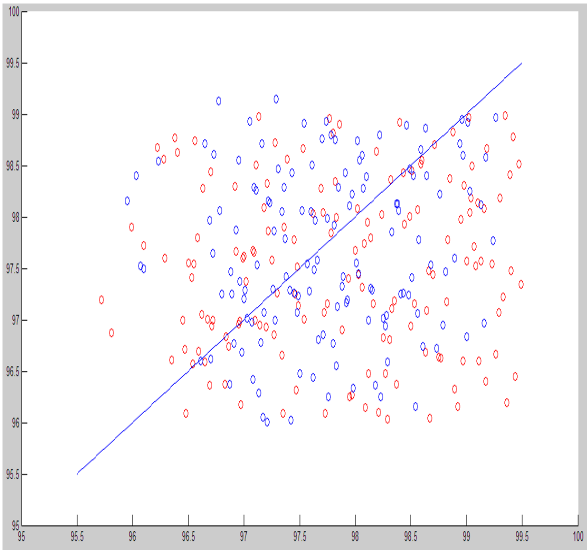

FIGURE IV. WL_DC_04 COMPARING PREDICTIVE RESULTS WITH ACTUALLY DETECTED DATA

Figure II, Figure III and Figure IV show little difference from Figure I, as the maximum difference between the actual values and predictive values is 4.0%, verifying the parameter predictive model is effective in utilization.

IV. SUMMARY

This paper first explains the necessity of data mining for desulfurization monitoring data, and then proposes to process the data using association rule in data mining algorithms. In association rule algorithms, in order to reduce the complexity of desulfurization monitor data analysis, data samples which are involved in desulfurization process will first go through reduction process. In this way, we can ignore the irrelevant factors and only focus on the important factors which affect the desulfurization facilities and the desulfurization efficiency. We have established the data model for the normal operation of the desulfurization facilities, and proposed parameters predictive model by analyzing and mining of the Power plant desulfurization monitoring data. And we analyze the feasibility and effectiveness of those models, and find the desulfurization efficiency is improved.

There are plenty of factors that can affect the operation of the desulfurization facilities, so if we want to analyze them accurately, we should clarify their meaning and characteristics, and the relationship between them. The data of this paper are obtained from a Shanghai Environmental Science and Technology Ltd. during its experiment and monitor on the plant desulphurization facilities. The capacity of the power plant desulphurization facilities is 600MW, and they use limestone gypsum wet to realize desulfurization treatment. In this paper we mainly considerate the fourteen factors of Table I which affect the desulfurization efficiency. If one changes the method of desulfurization processing or the size of the set, the desulfurization efficiency will change dramatically. The parameters predictive model we propose in this paper have a significant effect in intelligently managing for the power plant desulfurization facilities of this Shanghai Environmental Science and Technology Ltd. For other desulfurization facilities these two models may not be very consistent, but in those cases the methods and models of this paper can be adopted to obtain the corresponding parameter model. By studying the corresponding parameter model, one can provide accurate basis for environmental protection departments and power companies to make good decisions to achieve the common upgrade of social, economic and environmental benefits.

REFERENCES

[1] Yin Yuanhong, Li Honghua, Yang Jingjia. Intelligent environmental protection and technology application demonstration. <<Proceedings of the annual conference of the Chinese Academy of environmental science, 2012(Second volumes>>. 012-06-01.

[2] Wu Yong, Zhang Hongjian. Smart environmental solutions based on big data and cloud computing [J]. Information technology and standardization, 2013:Page 38-41

[3] Liu Rui, Zhang Zhiming, Xie Tao, Yao Xin, Sun Shiyou, Hou Litao. Discussion on the construction of "intelligent environmental protection" system in China [J]. Environmental protection and circular economy, 2012:9-14.

[4] Zhang Xinquan. The construction of intelligent environmental protection system and the case study of Xiangtan City [D]. Xiangtan: University Of Xiangtan, 2013:1-66.

[5] Xu Min,Sun Hailin. From "digital environmental protection" to ”intelligent environmental protection” [J]. Environmental monitoring management and technology, 2011, 23(4): 5-7.

[6] Parisa Rashidi, Student Member, IEEE,&Diane J. Cook, Fellow, IEEE&Lawrence B. Holder, Senior Member, IEEE&Maureen Schmitter-Edgecombe. Discovering Activities to Recognize and Track in a Smart Environment[J]. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, 2011, 23(4):527-539.

[7] Zhu Jianping, Zhang Runchu. The development and characteristics of data mining [J]. Statistics and decision making, 2002(7):71-72.

[8] G. Michael Youngblood, Diane J. Cook. Data mining for hierarchical model creation[J]. SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS,2007,37(4): 561-572.

[9] Zhou Changli, Guo Dongping. Development of flue gas desulphurization technology in coal fired power plant, China Coal [J], 2007(7).