CHEMICAL ENGINEERING TRANSACTIONS
VOL. 52, 2016

A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.aidic.it/cet

Guest Editors: Petar Sabev Varbanov, Peng-Yen Liew, Jun-Yow Yong, Jiří Jaromír Klemeš, Hon Loong Lam

ISBN 978-88-95608-42-6; **ISSN** 2283-9216
DOI: 10.3303/CET1652106

A Fuzzy Association Rule Mining Based Approach to Identifying SO_2 Emission Concentration Patterns of Coal-fired Boilers

Peng Ran*[a], Yase Wang[a], Pei Liu[b], Zheng Li[b]

[a]School of energy & power engineering, North China Electric Power University, Baoding, Hebei Province 071003, China
[b]State Key Lab of Power Systems, Department of Thermal Engineering, Tsinghua University, Beijing 100084, China
13582232866@163.com

Sulphur dioxide emissions control in coal-fired power plants faces significant challenges with increasingly stricter emissions standards. These challenges are further exaggerated by the fact that coal-fired power plants operate strongly irregularly in order to meet the user demand between full-load and part-load conditions, mainly due to requirements from a power grid with the increasing capacity of intermittent renewable power. Acquiring accurate SO_2 emissions patterns at these variable loads, thus unstable and unpredictable, are a challenging task. In this paper, we use a fuzzy association rule mining approach, based on real-life operational data, to obtain quantitative SO_2 emission concentration patterns of coal-fired boilers. Operational data from a 300 MW supercritical power plant over a one-year period are used for analysis purposes. Results show that an optimal operational strategy can be produced by applying the fuzzy association rule mining techniques. A coal-fired power plant, if operated following the optimal operation strategy, can reduce its coal consumption rate by 1.35 g/kWh and SO_2 emission concentration by 145 mg/Nm^3, respectively.

1. Introduction

In China, coal continues to play an important role as a primary fuel in energy consumption. Growing demand for electricity leads to further increase in the use of fossil fuels, like coals of different quality. Due to combustion of coals in large power plants significant amount of pollutant gasses, such as nitrogen and sulphur oxides, are released into the atmosphere. Environmental pollution has like-wise become a serious public and social issue in China. Most air pollution results from coal combustion and coal are the source of 90 % of the SO_2 emission, 70 % of the dust emission and 67 % of the NO_X emission (Liang et al., 2013). The Chinese government noticed this problem and begun to take measures to prevent further environmental deterioration. The standards for energy consumption and pollutant emission have been more rigorous, especially for coal-fired power plant's boilers (Wang and Feng., 2014). Therefore, it is required for coal-fired boilers to promote energy conservation and emission reduction. One target area to increase the energy efficiency of boilers and reduce pollutant emission is through investments in more efficient equipment and the use of high quality, low ash and low sulphur coal. It can be proven conclusively that improving coal quality contributes to an improvement in a boiler's environmental performance (Pan et al., 2014). Boiler efficiency and the quality of coals have a significant effect on environmental emissions, both gaseous and particulate and the lower the particulate emissions, the lower the concentration of trace elements released to the atmosphere (Yuan et al., 2013).

Compared to through investments way and the use of high quality coals, the operational parameters optimization poses a unique role for energy consumption and pollutant emission (Wang et al., 2012). When fuels containing sulphur compounds as contaminants are burned, substances such as sulphur dioxide, sulphur trioxide and sulphuric gas can be formed. The rate of formation of sulphur compounds in a combustion process can be alleviated by altering the operational parameter (Yuan et al.,2013), i.e. regulating the amount of oxygen available, well-coordinated primary air and secondary air, a good aerodynamic field in the furnace.

Please cite this article as: Ran P., Wang Y., Liu P., Li Z., 2016, A fuzzy association rule mining based approach to identifying so2 emission concentration patterns of coal-fired boilers, Chemical Engineering Transactions, 52, 631-636 DOI:10.3303/CET1652106

In addition, coal-fired power plant's energy consumption can be improved by optimizing the operational parameters according to the load of a boiler. However, the demands of power are time-sensitive, which often display a systematic variation over the hours of the day and over the seasons (Espatolero et al., 2014). This leads to the boiler working at variable load and deviating from the design-point, and requires more frequently manual effort because of coal varying greatly for a power plant in China. Besides, large scale coal-fired boiler is a complex nonlinear system with more uncertainties to describe the formation of sulphur compounds behaviour and to evaluate the SO_2 emission concentration for power generation, and is also difficult to present accurate dynamic mathematical models to predict the energy consumption of boiler operation (Mardon and Hower, 2004). So it is not easy to dynamically optimize the boiler practical operation in the traditional approach.

For every load and different kind of coals, a boiler has a different control strategy assigned to it, meaning a distinct pattern of pollutant emission. For coal making up 50 % of total fossil energy resources use for the power industry in China, there is lack of reliable estimates of how much energy can be saved and how much pollutant emissions can actually be reduced in this particular sector. To the best of the authors' knowledge, few published work on variations in SO_2 emission concentration patterns of coal combustion is available. Such information is essential for making innovative and emission reduction improvements in power plant's coal combustion system.

In view of the above discussions, an experimental research was carried out to investigate and quantify both the quantity and circumstances of coal combustion related SO_2 emission concentration of flue gas in a typical power plant at different loads over a full operational year. SO_2 emission concentration of flue gas and other operation characteristics data was gathered in a north China coal-fired power plant. A methodology based on fuzzy association rule mining (Han et al., 2012) was proposed to discover the most common patterns in controlling sulphur compounds emission and reducing the energy consumption. The discovered patterns were invoked as the operation optimization target values in a power plant operation optimization system to cut down the SO_2 emissions and lower the coal consumption.

The rest of this paper is structured as follows. A review of data association rule mining is described in Section 2. Then, in Section 3, the proposed methodology is presented. Case analysis and numerical results are provided in Section 4. Finally, Section 5 concludes this paper.

2. A review of data association rule mining

Data mining is the process of extracting valid, new and comprehensive information from huge amounts of data in order to improve and support business decisions (Hooke et al., 2013). In some cases, users may have no ideas regarding what kinds of patterns in their data may be interesting, and hence may like to search for several different kinds of patterns in parallel (Han et al., 2012). Thus, data mining requires verification and understanding of what the patterns means, to ensure that meaningful results are obtained from this process (Eldestein, 2012). Data association rule mining is a supervised method for discovering associations among a set of variables. Data association rule mining has served in a variety of applications such as energy (Cabrera and Zareipour, 2013), medicine (Delen et al., 2005) and marketing (Rathod and Garg, 2016).

Association rules describe events that have a tendency to occur together. They are formal statements in the form of X => Y, where if X happens, Y is likely to happen. Every association rule is composed by two different sets of items called an antecedent and a consequent. The antecedent and consequent frequently are called, respectively, the left-hand side (or LHS) and the right-hand side (or RHS). An association rule indicates an affinity between the LHS and the RHS. In order to select interesting rules from the set of all possible rules, constraints on various measures of significance and interest are used. The best-known constraints are minimum thresholds on support and confidence, and lift (Ye et al., 2003).

Data association rule mining algorithms search for rules that exceed user-specified threshold values of support, confidence and lift. In a naive data association algorithm, every possible combination of item sets is computed, evaluated and discarded based on the "lift" measures.

Let X be an item-set, X => Y an association rule and T a set of transactions of a given database. The support value of X with respect to T is defined as the proportion of transactions in the database which contains the item-set X, in formula: support(X). The confidence value of a rule, X => Y, with respect to a set of transactions T, is the proportion of the transactions that contains X which also contains Y. Confidence is defined as:

$$confidence(X \Rightarrow Y) = \frac{support(X \cup Y)}{support(X)} \tag{1}$$

The value of lift is that it considers both the confidence of the rule and the overall data set. The lift of a rule is defined as:

$$lift(X \Rightarrow Y) = \frac{support(X \cup Y)}{support(X) \times support(Y)} \quad (2)$$

To deal with the uncertainties and the quantitative values in real-world circumstance, researchers have started to hybridize fuzzy set theories into data mining. The fuzzy approach is used for transforming quantitative data into fuzzy data. It can make quantitative values have a smooth transition between fuzzy sets instead of using partition methods (Lee et al., 2015).Fuzzy association rule mining is one of the hybrid approaches taking the fuzziness of data into consideration. The knowledge discovered in terms of fuzzy association rules is believed to be more meaningful than that in terms of boolean association rules in real-life situations (Palacios et al., 2015).

The combustion process of coal-fired power unit is dense, especially when load rates are highly variable, when the coal quality varies, or when several interacting processes coexist at a single site (Ligang et al., 2014). Because of the load variation, dynamically of operation conditions, sensors measure and collect data in an imprecise way, a substantial number of uncertain and quantitative attribute data are transmitted, stored in databases, where data are numeric. Fuzzy sets not only provide soft boundaries but also conceive understandable semantics (Pei et al., 2013), and therefore fuzzy association rule mining fits well into the purpose of this study.

3. The proposed methodology for discovering SO_2 emission concentration patterns

The proposed methodology is presented in this section. Firstly, for the purpose of reducing noise and creating higher-level concepts, the raw data is pre-processed. Operation data normalization and attribute reduction are the core of the whole process. Operation data (x) are transformed to a value (x') between 0.0 and 1.0 by Min-Max normalization method. The lowest (min) value is set to 0.0 and the highest (max) value is set to 1.0. This provides an easy way to compare values that are measured using different scales or different units of measure.

$$x' = \frac{x - \min}{\max - \min} \quad (3)$$

Operation data of coal-fired power unit include the excess air coefficient of furnace, flue gas temperature, carbon content in fly ash, SO_2 emission concentration of flue gas, feed water temperature, main steam temperature, and so on. This type of parameter determines the operation characteristics of a power unit. On the other hand, to measure the efficiency of the boiler, coal quality and environment parameters are considered. All the historical data are stored in a data warehouse. All those parameters are identified for fuzzy association rule mining and then extracted from the data warehouse.

Before inputting these parameters into the mining algorithm in the knowledge discovery sector, the fuzzy characteristics of parameters have to be defined. Figure 1 represents an example of the normalization data and its membership functions. All fields have five fuzzy classes including: Very Low (VL),Low (L), Normal (N), High (H) and Very High (VH).These terms used for describing the quantitative values of parameters, and membership functions to consider the fuzziness of parameters.

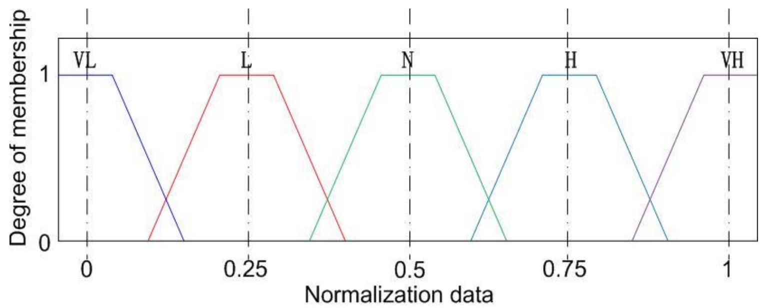

Figure 1: An example of the normalization data and its membership functions.

Secondly, an energy assessment is conducted in order to evaluate the efficiency of the boiler. Using the stored data, a preliminary energy assessment is conducted with the intent of quantifying coal consumption levels. The energy assessment is also allowed to identify when optimal operation conditions occur and their relationships with SO_2 emission concentration of flue gas in data preprocessing. By quantifying how much SO_2

emission concentration of flue gas there is and identifying when optimal operation conditions occur, potential efficiency opportunities are found. Energy assessments are important for comparison of SO_2 emission concentration of flue gas and energy performance under different scenarios and also provide long-term targets and supervision for guiding operation optimization.
Thirdly, support, confidence and lift thresholds are defined and a fuzzy data association mining algorithm is applied to the resulting database in order to discover knowledge and relationships between variables of the dataset with a focus on discovering the pattern of SO_2 emission concentration. For a rule to be considered strong, it must exceed the user-defined minimum thresholds for the aforementioned measurements. Apriori association mining algorithm (Han et al., 2012) is used to extract associations. Those associations are then evaluated using three measurements: support, confidence, and lift. The threshold for support and confidence depends on how many instances of the dataset the user wishes to cover. The minimum lift value is defined as 1, which means that for all associations the occurrence of the antecedent must imply the occurrence of the consequent (Han et al., 2012). The resulting association rules between SO_2 emission concentration patterns of flue gas and operation characteristics are analysed in order to extract knowledge from them and obtain operation optimization target values and SO_2 emission concentration reduction measures.

4. Case analysis and numerical results

4.1 Case unit and data collection system description

A 300 MW supercritical coal-fired power unit is selected as the case unit, which is located in north China. The boiler is of SG1080/17.60-M8 type, supercritical, intermediate reheat controlled circulation drum-type; the steam turbine is of N300-16.7/538/538 type, 300 MW supercritical condenser steam unit, single intermediate reheat and single-shaft.
The implemented data collection system is Supervised Information System (SIS), which is a set of real-time monitoring of the production process to optimize the operation and production process management as one of the plant-level automated information system, it is a real-time data acquisition and storage power plant DCS, plant-wide information sharing data plant form, and plant-wide real-time monitoring of the production process, calculation of economic indicators and analysis and other advanced application based on the database platform. Data on operation characteristics and SO_2 emission concentration of flue gas is automatically monitored every five seconds and sent over the intranet to a database server for storage.

4.2 Knowledge extraction and analysis

Before inputting those quantitative parameters into the mining algorithm, the fuzzy set theories were used to describe the quantitative values of parameters, and membership functions. The Apriori algorithm is chosen to complete the data mining process (Ran et al., 2012). The objective function is to maximum boiler efficiency and minimum SO_2 emission concentration of flue gas.
There are four types of coal for the coal-fired power plant, the elemental compositions of different coals are shown in Table 1. Under the conditions of 100 %, 75 % and 50 % power load, by using the type a coal, and the knowledge extraction results are shown in Table 2.
It can be seen from the data in Table 2 that the boiler efficiency η_b reaches maximum value (no more than 92 %) and the SO_2 emission concentration of flue gas reaches maximum value (no more than 1800 mg•Nm^{-3}) when the excess air coefficient α is set between 1.21 and 1.25 at full load range (296 MW~300 MW), and the SO_2 emission concentration of flue gas increases with a decrease in the excess air coefficient α.
For the purpose of analyzing the associations relationship between the coal quality and SO_2 emission concentration of flue gas, rules of SO_2 emission concentration of flue gas by burning different kind of coal were extracted, which are shown in Table 3. The Table 3 illustrates the SO_2 emission concentration of flue gas increases with a rise in sulphur content.

Table 1: Ultimate analytical data of four selected coal

Coal type	C_{ar}(%)	H_{ar}(%)	O_{ar}(%)	N_{ar}(%)	S_{ar}(%)	A_{ar}(%)	M_{ar}(%)	$Q_{ar,net}$(kJ/kg)
a	67.92	4.13	7.59	0.76	0.73	10.12	8.75	26,583
b	57.12	3.61	9.22	0.96	0.49	12.58	16.02	22,362
c	61.22	3.24	7.52	0.45	1.10	18.27	8.20	23,189
d	50.72	3.12	7.33	0.69	1.56	29.99	6.59	19,281

Table 2: Rules of typical load by using the coal type a

Load	Rules of typical load	Support	Confidence	Lift
100 %	LOAD ∈ (296~300 MW) & η_b ∈ (91.38~91.65 %) & α ∈ (1.21~1.25) => SO_2 (1,679~1,745 mg•Nm^{-3})	0.401	0.852	1.58
75 %	LOAD ∈ (221~225 MW) & η_b ∈ (91.35~91.61 %) & α ∈ (1.29~1.31) => SO_2 (1,608~1,587 mg•Nm^{-3})	0.395	0.844	1.62
50 %	LOAD ∈ (147~150 MW) & η_b ∈ (90.21~90.43 %) & α ∈ (1.38~1.41) => SO_2 (1,553~1,491 mg•Nm^{-3})	0.388	0.846	1.57

Table 3: Rules of 100 % load by using different kind of coal

Coal type	Rules of typical load	support	confidence	lift
a	LOAD ∈ (296~300 MW) & η_b ∈ (91.38~91.65 %) & α ∈ (1.21~1.25) => SO_2 (1,679~1,745 mg•Nm^{-3})	0.403	0.852	1.58
b	LOAD ∈ (296~300 MW) & ηb ∈ (91.33~91.62 %) & α ∈ (1.21~1.25) => SO_2 (1,386~1,448 mg•Nm^{-3})	0.421	0.835	1.58
c	LOAD ∈ (296~300 MW) & η_b ∈ (91.35~91.63 %) & α ∈ (1.21~1.25) => SO_2 (2,936~3,060 mg•Nm^{-3})	0.401	0.845	1.61
d	LOAD ∈ (296~300 MW) & η_b ∈ (91.31~91.62 %) & α ∈ (1.21~1.25) => SO_2 (4,943~5,156 mg•Nm^{-3})	0.392	0.851	1.57

In addition, middle values or weighted average of those intervals can be adopted as the operation optimization target values. For instance, the operation optimization target value of excess air coefficient α is 1.23 when the load is 300 MW (100 %) by utilising the weighted average method. In a given typical load range, the excess air coefficient related to low coal consumption and low SO_2 emission concentration of flue gas is chosen as operation optimization target value to guide the operation. The operation optimization target values obtained from fuzzy association rule mining and designed values are listed in Table 4.

Table 4: Operation optimization target value (Excess air coefficient)

LOAD	300 (MW)	225 (MW)	150 (MW)
Design value	1.21	1.33	1.42
Optimization target values	1.23	1.30	1.395

4.3 Operation optimization

The operation optimization target values obtained by fuzzy association rule mining from the operation data are more suitable for guiding operators to achieve an efficient performance by adjusting the controllable parameters such as the excess air coefficient of furnace, coal quality, the main steam temperature, main steam pressure, reheated steam temperature, and so on. The process parameters are optimized based on the results of fuzzy association rule mining to examine the effects of the optimization in a north China coal-fired power plant. The average boiler efficiency improved about 0.26 %. The coal consumption reduced about 1.35 g/kWh and the SO_2 emission concentration of flue gas decrease about 145 mg•Nm^{-3}.

5. Conclusion

In this work, a fuzzy data association rule mining-based method is proposed to uncover possible relationships between SO_2 emission concentration patterns and operation characteristics. The relationships between SO_2 emission concentration of flue gas and the excess air coefficient of furnace, coal quality and power unit load were discussed. The SO_2 emission concentration of flue gas increases with a rise in sulphur content, and the SO_2 emission concentration of flue gas increases with a decrease in the excess air coefficient α. The operation optimization target values obtained by this work are more suitable for guiding operators to achieve an efficient performance by adjusting the controllable parameters.

The two major contributions of this paper can be summarized as: (i) to investigate the SO_2 emission concentration of flue gas patterns and their correlation with operation patterns in coal-fired power plant based on high resolution real-life data and (ii) to apply fuzzy data association mining for SO_2 emissions analysis and discovering SO_2 emission concentration patterns in easy-to-interpret rules. The significance of the first contribution is that it fills a gap in the availability of realistic estimates and understanding of both SO_2 emission concentration and coal consumption in coal-fired power unit. The second major contribution is significant because it introduces a powerful data mining tool for investigating SO_2 emission concentration patterns.

Acknowledgments

This study was supported by National Natural Science Foundation of China (Grant No. 51506052), and the Fundamental Research Funds for the Central Universities (No. 2014MS144).

References

Cabrera D. F. M., Zareipour H., 2013, Data association mining for identifying lighting energy waste patterns in educational institutes, Energy and Buildings, 62, 210-216

Delen D., Walker G., Kadam A., 2005, Predicting breast cancer survivability: a comparison of three data mining methods, Artificial Intelligence in Medicine, 34, 113-127

Eldestein H. A., 1999, Introduction to Data Mining and Knowledge Discovery, Two Crows Corporation, Potomac, MD, USA

Espatolero S., Romeo L. M., Corts C., 2014, Efficiency improvement strategies for the feedwater heaters network designing in supercritical coal-fired power plants, Applied Thermal Engineering, 73, 447-458.

Han J., Kamber M., 2012, Data Mining: Concepts and Techniques. Elsevier, Beijing, China.

Hooke J. H., Landry B. J., Eng P., David H. M., 2003, Energy management information systems: achieving improved energy efficiency: a handbook for managers. Office of Energy Efficiency of Natural Resources Canada, Ottawa, Canada.

Lee C. K. H., Tse Y. S., Ho G. T. S., Choy, K. L., 2015, Fuzzy association rule mining for fashion product development, Industrial Management & Data Systems, 115, 383-399.

Liang X. Y., Wang Z. H., Zhou Z. J., Huang Z. Y., Zhou J. H., Cen K. F., 2013, Up-to-date life cycle assessment and comparison study of clean coal power generation technologies in China, Journal Of Cleaner Production, 39, 24-31.

Ligang W., Yongping Y., Changqing D., Morosuk T., Tsatsaronis G., 2014, Parametric optimization of supercritical coal-fired power plants by MINLP and differential evolution, Energy Conversion and Management, 85, 828-38.

Mardon S. M., Hower J. C., 2004, Impact of coal properties on coal combustion by-product quality: examples from a Kentucky power plant, International Journal of Coal Geology, 59, 153-169.

Palacios A. M., Palacios J. L., Sanchez L., Alcala-Fdez J., 2015, Genetic learning of the membership functions for mining fuzzy association rules from low quality data, Information Sciences, 295, 358-378.

Pan J. N., Kung P., Bretholt A., Lu J. X., 2014, Prediction of energy's environmental impact using a three-variable time series model, Expert Systems with Applications, 41, 1031-1040.

Pei B., Zhao S. Y., Chen H., Zhou X., Chen D. J., 2013, FARP: Mining fuzzy association rules from a probabilistic quantitative database, Information Sciences, 237, 242-260.

Ran P., Li G. S., Zhang S. F., Wang S. L., 2012, Decision of the most economical coal for power plants based on data mining technology, Journal of Chinese Society of Power Engineering, 32,494-500 (in Chinese).

Rathod R. R., Garg R. D., 2016, Regional electricity consumption analysis for consumers using data mining techniques and consumer meter reading data, International Journal of Electrical Power & Energy Systems, 78, 368-374.

Wang N., Zhang Y., Zhang T., Yang Y., 2012, Data Mining-Based Operation Optimization of Large Coal-Fired Power Plants, AASRI Procedia, 3, 607-612.

Wang Z. H., Feng C., 2014, The impact and economic cost of environmental regulation on energy utilization in China, Applied Economics, 46, 3362-3376.

Xu J., Yang Y., Li Y. W., 2015, Recent development in converting coal to clean fuels in China, Fuel, 152, 122-130.

Ye N., 2003, The Handbook of Data Mining, Lawrence Erlbaum Associates,Mahwah, New Jersey, United States.

Yuan X. L., Mi M., Mu R. M., Zuo J., 2013, Strategic route map of sulphur dioxide reduction in China, Energy Policy, 60, 844-851.