Available online at www.sciencedirect.com

SciVerse ScienceDirect

AASRI Procedia 1 (2012) 142 – 149

AASRI Procedia

www.elsevier.com/locate/procedia

2012 AASRI Conference on Computational Intelligence and Bioinformatics

The Extraction Method of the Energy Consumption Characteristics Based on Fuzzy Rough Set

Zhao Ming [1], Yan Zhengbo [1], Zhou Liukun[1], Wang Huijie [2],Xu Xiaogang[2]

[1]Yunnan Electric Power Test & Research Institute Group Co. Ltd. Electric Power Research Institute, Kunming 650217,Yunnan Province, PRC
[2]North China Electric Power University, Baoding 071003
E-MAIL: zhouvic888@163.com

Abstract

The large coal-fired power generation system characterizes as complex structure, highly coupling and nonlinear correlation between energy consumption and external environment, resources and load demand. The attribute reduction method based on fuzzy rough set (FRS) was introduced to extract the dominant features related to energy consumption and eliminate the redundant features as well. Taking the reduced attributes as input variables, which provide the important base for analyze the energy consumption of the coal-fired power unit.

Keywords: Energy consumption; associative relation; fuzzy rough set; reduction rule; data mining.

1. Introduction

The large coal-fired power generation system was composed of many complex subsystems, and there were high-dimensional nonlinear associative relationships among various operating parameters. To determine the relationships among parameters under different operating conditions, and reconstruct the system operating state, thus the optimum controllable parameters of the boundary condition can be obtained, and then the operation optimization strategies can be formulated. All of this can be summarized as the analysis of complex systems modeling.

Due to the characteristics of large span thermodynamic state, large flow, high hear flow density and large scale equipments etc, the nonlinear scale effect on energy transport, conversion and quality dissipation in

doi:10.1016/j.aasri.2012.06.023

different levels of systems, processes and unit equipments of the large coal-fired generating units was obvious. At the same time, there were strong relationships between the power generation energy consumption and the environment, resources, loads.

As the main way of energy production, large-scale coal-fired units, including (ultra) supercritical units, will bear rapid peak regulation operation, frequent change of working conditions in long time, stand the test of factors such as complex coal quality, the cold junction and environment etc. Its energy consumption level has distinct time-varying characteristics. Considering the characteristics of large coal-fired generating units, as well as the coupling effects of external factors, it was an important foundation of achieving energy saving of coal-fired power generation process to reveal the spatial and temporal distribution of unit energy consumption within the scope of total working conditions. However, due to the characteristics such as mechanism complexity, nonlinear, time-varying, large time delay, strong coupling, uncertainties and random disturbances of the equipment, systems and process, the traditional mechanism modeling method was difficult to establish the exact energy consumption space-time distribution model.

The data in unit generating process database can reflect the operating conditions and characteristics of operating variables, process equipment and process directly or indirectly, and the data mining technology provides a powerful tool for knowledge discovery in databases. From the vast amounts of data, it can dig out useful information and knowledge, thus the powerful decision supports for the energy consumption online monitoring of large-scale coal-fired units, optimization of energy-saving operation and management can be provide.

In response to these characteristics, features were extracted from real data of thermal power units though attribute reduction method based on fuzzy rough set (FRS) in basis of data mining technology in the paper. The relationship was analyzed among variables, between variables and energy consumption distributes, to reduce attribute table, drop redundant data. That provides the important research base for real unit energy consumption characteristics analysis in depth.

2. Fuzzy rough sets theory

Classical rough set theory proposed by Pawlak, was used to depict the incompleteness and uncertainty. It can classify information which was inaccurate, inconsistent and incomplete for effectively, and discover hidden knowledge, reveal the potential law.

The function of knowledge classification of the rough set theory was achieved through attribute reduction. Attributes (characteristics) of the information system was not equally important, or even some of the attributes was redundant. The system will be mostly simplified by reducing attribute of knowledge in maintaining the system's ability of classification, deleting irrelevant or unimportant attribute knowledge to extract the most relative characteristics and variation of the properties.

The following was the relevant definition of the reduction of knowledge in rough set theory. Information system $S=<U, C\cup D, V, f>$, U was a non-empty objects set; C and D were the properties sets. Namely the condition attribute set was C and decision attribute set was D; V was a set of attribute value ; f was an information function. namely $f: U\times(C\cup D)\rightarrow V$. It specifies the object's property values of U in each group.

In the given knowledge base K= (U, S), U was universe of discourse, S was the equivalence relation clusters on U. Then for each subset $X\subseteq U$ and equivalence relation $R\in IND(K)$,

$$R^*(X) = \cup \{x \in U : [x]R \subseteq X\} \tag{1}$$

X was R's lower approximation.

The dependence of two attribute sets between C and D was defined as

$$\gamma C(D)=|POSC(D)|/|U| \tag{2}$$

POSC(D)=C*(D)；|U| represents the entire collection of the number of objects.
For the condition attributes ai∈C, ai on the importance of D was defined as

$$\gamma ai\ (D)= \gamma C(D)- \gamma C\text{-}\{ai\}\ (D) \tag{3}$$

The classical rough set theory can only deal with discrete or symbol-based property information system. The large-scale coal-fired unit energy consumption decision table (Table 2) based on the unit's actual operating data (shown in Table 1) was continuous real-value in this paper, and they need to be reduced by attribute discretization methods. But the discretization process will bound to lose some important information to have a negative impact on the classification results. Therefore, the reduction method based on fuzzy rough set theory attributes were adopted in the paper, fuzzy sets were used to replace the precise set, the fuzzy similarity relation was introduced to instead equivalence relations on the universe of discourse, the classical rough set theory was extended to fuzzy rough sets.

Table 1. Actual operating data of large-scale coal-fired units to indicate

Load (MW)	Main steam pressure (Pa)	Main steam temperature (℃)	…	Reheat steam temperature (℃)	Circulating water inlet temperature (℃)	Circulating water flow (kg/h)	Coal consumption (g/kwh)
460.6746	16.1340	533.6404	…	518.5391	14.9309	39865.31	312.0035
460.75	16.1041	533.6458	…	518.2047	14.9815	39904.81	312.1624
460.6969	16.0613	533.3089	…	517.7281	14.9865	39921.98	312.26
460.134	16.0223	532.6219	…	516.8508	14.9723	40072.98	312.4532
460.0841	15.9722	531.6393	…	515.9906	14.9365	40155.88	312.7877
459.6624	15.9278	530.5906	…	515.3385	14.9331	40359.6	313.3375
460.2422	15.8903	529.6125	…	515.0047	14.9124	40408.05	313.4895
460.0361	15.8641	528.8219	…	515.0008	14.9227	40539.27	313.5216
459.3332	15.8385	528.3969	…	514.8586	14.9464	40575.51	313.7807

Table 2. Large coal-fired unit energy consumption decision-making table

Sample	Condition attributes (C)							Decision attribute
	C1	C1	C2	C3	C4	…	Ck	D
X1	460.6746	16.1340	533.6404	14.9309	518.5391	…	39865.31	335.0035
X2	460.75	16.1041	533.6458	14.9815	518.2047	…	39904.81	335.1624
X2	460.6969	16.0613	533.3089	14.9865	517.7281	…	39921.98	335.26
X3	460.134	16.0223	532.6219	14.9723	516.8508	…	40072.98	335.4532
X4	460.0841	15.9722	531.6393	14.9365	515.9906	…	40155.88	335.7877

X5	459.6624	15.9278	530.5906	14.9331	515.3385	…	40359.6	336.3375
X6	460.2422	15.8903	529.6125	14.9124	515.0047	…	40408.05	336.4895
X7	460.0361	15.8641	528.8219	14.9227	515.0008	…	40539.27	336.5216
X8	459.3332	15.8385	528.3969	14.9464	514.8586	…	40575.51	336.7807
…	…	…	…	…	…	…	…	…

The following was relevant definitions of attribute reduction in f fuzzy rough set. Fuzzy relations RB was defined as

$$RB = \{(x,y) \in U \times U : \mu_{R_B}(x , y) \} \quad (4)$$

Where: RB∈F(U×U); μ_{R_B} (x , y)∈[0,1] the size of the (x,y) reflects the degree of (x,y) belonging to RB. If\ satisfied: 1) symmetry: μ_{R_B} (x , y) = μ_{R_B} (y, x) , ∀ x,y∈U; 2）Reflexive: μ_{R_B} (x , x) =1， ∀ x∈U， namin RB was a fuzzy similarity relation on U.
For ∀ x∈U， define

$$R_B^\lambda = \left\{ y \in U : \forall \mu_{R_B}\left(y,x\right) \geq \lambda, yR_B^\lambda x \right\} \quad (5)$$

Where: λ was the given threshold, R_B^λ was a collection of objects whose similarity of x over λ. Basing on the fuzzy similarity relation X, the lower approximation R_B^λ was defined as

$$R_{B*}^\lambda(X) = \cup \left\{ x \in X : R_B^\lambda\left(x\right) \subseteq X \right\} \quad (6)$$

Positive region of R_B^λ of X was defined as

$$POS_{R_B^\lambda}(X) = R_{B*}^\lambda(X) \quad (7)$$

the dependence γC(D) of two attribute sets between C and D was defined as

$$\gamma C(D) = |POSC(D)|/|U| \quad (8)$$

To find out the importance of certain attributes, the heuristic algorithm was adopted. Properties were added one by one, and then changes investigated. If the corresponding changes were great after adding attribute in the classification, that was high important; on the contrary, that was low.

The importance of attribute Ci on D was defined as:

$$\gamma ai\ (D)= \gamma C+\{ai\}\ (D)-\gamma C(D) \quad (9)$$

γC+{ai } (D) was the dependence of condition attribute relative to decision attribute after adding attribute ai,. If γai (D) = 0, so the attribute ai can be removed from the attribute set. It can be obtained a relationship between the condition attributes and decision attributes, and the set of {Ci} which plays a decisive role in the properties of the decision-making for D can be extracted , namely the energy consumption characteristics (referred to here as ‘energy label’) can be extracted.

3. Attribute reduction algorithm based on fuzzy rough set

The data in the paper was collected from the panshan power plant, Unit No. 3, 3250 data points total from March 2006 to May 2006. The 2260 controllable sample points were got though quasi-steady-state test, error and redundant data elimination, and other data de-noising and cleaning process. The energy consumption of a decision table was shown in Table 2. The decision table consists of 54 condition attributes (including controllable variables and uncontrollable variables) and one decision attribute (select supply coal consumption for decision-making attributes). The date of the unit energy consumption decision-making table, in order to eliminate the dimensionless impact, was normalized such as the type of the data shown in (10).

$$z_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (10)$$

x_i and z_i respectively were observation variables before and after normalization. $x_{\max}$ and $x_{\min}$ respectively were the maximum and minimum values of the observed value of variable x_i.

Simple algorithm for the basic process was as follows:

1. Computing every $DIS(R)$ and $DIS(\mathbf{R})$;
2. Select one $R \in \mathbf{R}$;
3. Select $(x_{i_0}, x_{j_0}) \in DIS(R)$, compute $\bigcup\{DIS(R') : (x_{i_0}, x_{j_0}) \notin DIS(R')\}$ and $\bigcap\{DIS(R_0) : (x_{i_0}, x_{j_0}) \in DIS(R_0)\}$;
4. Compute $DIS(\mathbf{R}) - \bigcap\{DIS(R_0) : (x_{i_0}, x_{j_0}) \in DIS(R)\}$;
5. If $DIS(\mathbf{R}) - \bigcap\{DIS(R_0) : (x_{i_0}, x_{j_0}) \in DIS(R_0)\} \subseteq \bigcup\{DIS(R') : (x_{i_0}, x_{j_0}) \notin DIS(R')\}$, go to step 8;
6. If $DIS(\mathbf{R}) - \bigcap\{DIS(R_0) : (x_{i_0}, x_{j_0}) \in DIS(R_0)\} \subseteq \bigcup\{DIS(R') : (x_{i_0}, x_{j_0}) \notin DIS(R')\}$ does not hold, let $DIS(R) = DIS(R) - \{(x_{i_0}, x_{j_0})\}$ and go to step 3;
7. If $DIS(R) = \phi$, go to step 2;
8. Let $SLECTION = SELECTION \bigcup \{(x_{i_0}, x_{j_0})\}$ and $\mathbf{R} = \mathbf{R} - \{R\}$, then go to step 2;
9. Output $SLECTION$ when $\mathbf{R} = \phi$;
10. Compute $\{c_{ij} : (x_i, x_j) \in SLECTION\}$;
11. Let $REDUCT = \{R : c_{ij} = \{R\}\}$ and $c = \{c_{ij} : (x_i, x_j) \in SLECTION\} - \{R : c_{ij} = \{R\}\}$;
12. Select the most frequently $R \in c_{ij} \in c$;
13. Let $REDUCT = REDUCT \bigcup \{R\}$ and $c = c - \{c_{ij} : R \in c_{ij} \in c\}$,
    go to step 12;
14. Stop and output $REDUCT$ when $c = \phi$.

Figure1 based on fuzzy rough set attribute reduction algorithm processes

Similar relationship matrix among variables was get from the similar relationship, the part of matrix was listed below:

$$R_B = \begin{bmatrix} 1 & 0.8696 & 0.7547 & 0.6957 & 0.6087 & 0.6957 & 0.6087 & 0.5683 & 0.587 & 0.7826 \\ & 1 & 0.7516 & 0.8261 & 0.7391 & 0.6926 & 0.7391 & 0.8137 & 0.7174 & 0.795 \\ & & 1 & 0.7174 & 0.7391 & 0.8261 & 0.7391 & 0.8137 & 0.7174 & 0.8356 \\ & & & 1 & 0.8106 & 0.6584 & 0.7516 & 0.5311 & 0.8913 & 0.7609 \\ & & & & 1 & 0.8478 & 0.925 & 0.7205 & 0.8913 & 0.8261 \\ & & & & & 1 & 0.9068 & 0.8727 & 0.7391 & 0.8975 \\ & & & & & & 1 & 0.7795 & 0.8323 & 0.8261 \\ & & & & & & & 1 & 0.6118 & 0.7702 \\ & & & & & & & & 1 & 0.8044 \\ & & & & & & & & & 1 \end{bmatrix} \tag{11}$$

Specific reduction results as follows:

global red={[1 2 3 10 4 9 7 6 5 8]，[1 2 4 5 8 10 9 7 6 3]，[1 2 3 4 6 7 8 9 5 10]，[7 9 10 1 6 5 2 4 3 8]，[1 2 3 4 5 6 7 9 10 8]，[1 2 4 5 6 7 9 10 3 8]，[1 2 4 5 6 8 9 10 7 3]，[1 2 3 4 5 6 8 9 10 7]}；

local_red={[1 2 3 10 4]，[1 2 4 5 8 10]，[1 2 3 4 6 7 8 9]，[7 9]，[1 2 3 4 5 6 7 9 10]，[1 2 4 5 6 7 9]，[1 2 4 5 6 8 9 10]，[1 2 3 4 5 6 8 9 10]}；

The supply coal consumption of decision attribute was discrete fuzzy, the class was divided into 8 types. Divide universe of discourse in the T-similarity relation. Global about and partial reduction get the 10 key energy consumption characteristics (Table 3) were calculated based on fuzzy rough set attribute reduction algorithm shown in Figure 1.

Table 3. Key energy consumption feature extraction results

1	2	3	4	5	6	7	8	9	10
Load	Main steam pressure	Main steam temperature	Reheat steam temperature	Circulating water inlet temperature	Reheat to reduce the temperature flow	Furnace export oxygen	Exhaust gas oxygen	Exhaust gas temperature	Circulating water
MW	MPa	℃	℃	℃	t/h	ppm	ppm	℃	t/h

4. Theoretical verification of the results

Plant system was a complex thermal system composed of many subsystems (equipment) which connected by specific way, working under multi-boundary conditions, completing the process of energy conversion thermal energy to mechanical energy and ultimately to electricity.

Thermal system with specific device structures, material properties and system structures conducts the process of combustion, heat transfer, mass transfer and flow through each device inside and shows corresponding running state parameters in the constrained conditions such as uncontrollable boundary conditions (such as power generation load, the local meteorological conditions, coal quality, etc.), controllable boundary conditions (such as the turbine initial operating parameter, circulating water flow, boiler primary or second air flow, etc.),system structures and device characteristics. These state parameters were ultimately expressed as the thermal and economic performance indicators of the unit’s thermal efficiency or power consumption rate.

The most fundamental indicators of energy performance of thermal power units usually expressed as the coal consumption rate b_{sn} :

$$b_{sn}=\frac{123}{\eta_b\eta_i\eta_m\eta_g\eta_p(1-\sum\xi_i)} \tag{12}$$

Where, $\eta_b,\eta_i,\eta_m,\eta_g,\eta_p,\sum\xi_i$ respectively were boiler efficiency, cycle thermal efficiency, mechanical efficiency, generator efficiency, pipeline efficiency and auxiliaries electricity consumption rate.

To specific unit, equipment structure, material properties and system structure, even specific defects have been fixed. So b_{sn} can be expressed by function of system boundary conditions:

$$b_{sn} = f(N_g, T_{xrw}, D_w, P_0, T_0, T_{rh}, C_{coal}) \tag{13}$$

Where, $N_g, T_{xrw}, D_w, P_0, T_0, T_{rh}, C_{coal}$ respectively were load, circulating water inlet temperature (determined by the cooling tower performance and the ambient temperature), the circulating water flow, main steam pressure, main steam temperature, reheat steam temperature, coal characteristics.

Therefore, the 10 attributes based on fuzzy rough set attribute reduction algorithm were extracted as input variables. The conclusions were consistent with the theoretical analysis; the key parameters affecting the unit energy consumption were the unit boundary condition parameters for the operation. This method provides an important basis for further study on optimization of unit consumption characteristics and operating parameters.

References

[1]. Diyong Jie, Wang Guopeng, Han Pu, system identification based on support vector machines [J]. Computer simulation, 2004,21 (11), 39 ~ 41.
[2]. Wu Hongxiao, PROCEEDINGS, Liu Yong, et al. Immune clustering-based radial basis function network model of short-term load forecasting [J]. Chinese Society for Electrical Engineering, 2005,25 (16): 53 ~ 56.
[3] Chen Jian Hung, Li Wei, Sheng Deren. Line performance calculation of a thermal power units in data fusion [J]. Chinese Society for Electrical Engineering, 2002,22 (5) :152-156.
[4]. Wang Huijie thermal power units operating parameters of energy consumption sensitivity analysis [J]. Chinese Society for Electrical Engineering, 2008,28 (29) :6-10.
[5]. Tat Yan, Hu Qinghua, Bowen. Integration of rough sets and fuzzy clustering, continuous data knowledge discovery [J]. Electrical Engineering of .2004,24 (6) :205-210.
[6]. JWHines, REUhrig, DJWrest. Use of Autoassociative Neural Network for Signal Validation [J]. Journal of Intelligent and Robotic Systems, 1997.143.
[7] Zhou Shuxiao Based on Wavelet Denoising and neural network theory of gas-solid circulating fluidized bed particle concentration prediction [J]. Petrochemical .2003,32 (3): 224 ~ 229. [J]. Chinese Society for Electrical Engineering .2004,24 (7) :253-256.
[8]. The mathematical model of the virtual sensors in Bo, Hu Nien Su, Zhou Yuyang. Steam Turbine Monitoring and Diagnosis System [J]. China Electrical Engineering .2004,24 (7) :253-256.
[9] SI Feng-qi, XU Zhi-gao. Based on the self-associative neural network and the measurement data since the calibration test method [J]. Chinese Society for Electrical Engineering .2002,22 (6) :153-155.
[10] Liu Fuguo. Power plant coal into the furnace element analysis and heat of the soft sensor real-time monitoring [J]. Electrical Engineering of .2005,25 (6) :139-146.
[11]. Zhang Zhiming, Chenghui Tao, Hong Xu, Hu, three high in the neural network ensemble forecasting model and its hair in the steam turbine unit state of repair [J]. Electrical Engineering of .2003,23 (9): 204 ~ 206.
[12] Chen Huayou combination forecasting method and its application: [doctoral thesis], University of Science and Technology, 2002,9.
[13]. Dongxue Yu, Hu China, Xu Zhi-gao. Power plant performance analysis of sampled data reliability test method [J]. Power Engineering .1998,18 (2): 16 to 19,74.
[14] Li Jianqiang. Optimization theory and applications of data mining-based power plant operation: [doctoral thesis], North China Electric Power University, 2006.
[15]Toda, Hiromichi, Yamanaka, "Planning and operation performance of 600 MW coal and oil dual-fired boiler for a supercritical sliding pressure operation" [J]. Technical Review - Mitsubishi Heavy Industries, Vol 22, No. 3, pp.225-233, Oct, 1985.