

Research on Energy Management Strategy of Captive Power Plant Based on Hybrid Method

Juan Liu, Xixia Huang,* Xiaoli Liu, and Zhiliang Han

Key Laboratory of Marine Technology and Control Engineering Ministry of Communications,
Shanghai Maritime University, No. 1550, Lingang Road, Pudong New Area, Shanghai 201306, China

(Received January 4, 2019; accepted June 3, 2019)

Keywords: captive power plant, energy management strategy, unit energy consumption, RF-RFE, optimization

Compared with a regular power plant, the load and fuel composition of a captive power plant vary greatly, resulting in complex and variable working conditions, and for specific operating conditions, there is a lack of data. Considering the irregularity and incompleteness of the data, a data-mining-based boiler energy management analysis method was proposed in this paper. This study adopted the random forest and recursive feature elimination (RF-RFE) method to select the features of distributed control system (DCS) monitoring data. According to different load intervals, some data with better running conditions were selected for clustering. According to the clustering center, adjustable parameters were optimized, and the obtained optimization results were input to the support vector machine model for energy consumption comparison. The results show that the method used in this study can reduce the unit energy consumption by 3–5% by adjusting the controllable parameters.

1. Introduction

Traditionally, a coal-fired self-contained power plant is an important part of the thermal power industry. It provides power supply for the production and operation of industrial enterprises and reduces the production cost of enterprises. At the same time, it also takes into account the electricity and heat demand of surrounding enterprises and residents. Some cities have problems such as the large capacity of the installed power grid, the low public load, the high proportion of coal-fired heating and self-sufficient power plants, and the limited new energy. With the wide applications of the supervision information system (SIS) and distributed control system (DCS) in power plants, data mining technology has been widely used in the optimization of power plant management units. Self-owned power plants have problems such as poor data conditions, large load variations (20–100%), and unstable materials. As a result, there is a problem of insufficient data for specific working conditions. Therefore, it is of great practical significance to adopt appropriate data screening and analysis methods.

At present, the energy consumption management of a power plant aims at energy saving, emission reduction, and prolonging the service life of equipment. Boilers are the key equipment

*Corresponding author: e-mail: huangxx@shmtu.edu.cn
<https://doi.org/10.18494/SAM.2019.2314>

for energy conversion. The energy consumption management of a power plant mainly focuses on the modeling of a boiler combustion system based on big data,^(1–13) the optimization of a combustion coal mixture strategy,^(14–16) and boiler unit equipment improvement.⁽¹⁷⁾ References 1–5 indicate the use of the neural network method to model the key parameters of boiler combustion optimization and the selection of input parameters that is only based on manual operation experience. Kim *et al.*⁽⁶⁾ studied the estimation of NO_x emission from coal-fired boilers of 500 MW units and the selection of input parameters. On the basis of auto regressive moving average (ARMA), artificial neural network (ANN), and least squares support vector machine (LSSVM) modeling methods, the appropriate input parameters of the NO_x emission model were selected by a sensitivity analysis method. Cheng *et al.*⁽⁷⁾ built a multiobjective optimization system for the energy consumption of coal-fired boilers. Rahat *et al.*⁽⁸⁾ used the data-driven Gaussian process (GP) regression model to establish the NO_x emission and carbon content model of fly ash. Then, they used the evolutionary multiobjective search algorithm to optimize the model parameters and gave the balance scheme of NO_x emission and boiler efficiency under different loads. Wang *et al.*⁽⁹⁾ studied the application of the GP in the optimized combustion process of reducing the NO_x emission of 330 MW boilers and applied the GP to simulate the relationship between NO_x emission characteristics and boiler operating parameters. The genetic algorithm (Ga) was used to optimize the GP model. The case study showed that the method can adapt to changes in working conditions and the prediction of adaptive learning is effective. Wang *et al.*⁽¹⁰⁾ studied a 600 MW subcritical twin-wall double-flue coal-fired boiler and established four models (BPNN, PCABPNN, PLSBPNN, and DBNBPNN) based on a reverse neural network (BPNN) to predict NO_x emissions of coal-fired power plants. The feature selection method based on information interaction was used to optimize the input features of the model. The results show that the model based on the DBN had a higher prediction accuracy and a higher stability.

As indicated above, different data mining methods were used to model the boiler efficiency and analyze the operable parameters, aiming at reducing the boiler energy consumption and pollutant emission. However, the boilers studied were the energy conversion equipment of regular power plants with management specifications and good data conditions. At present, there are few related studies on captive power plants. Taking a self-supplied power plant as the research object, considering the irregularity and incompleteness of the data, a data-mining-based boiler energy management analysis method is proposed in this paper.

The rest of this paper is structured as follows. Section 2 shows the relevant theoretical basic methods, Sect. 3 introduces the energy management methods of a captive power plant, and Sect. 4 describes the experiment and shows experimental results and finally conclusions.

2. Theoretical Basis

2.1 Random Forest and Recursive Feature Elimination (RF-RFE) Algorithm

The recursive elimination feature method uses a machine learning model to perform multiple rounds of training.⁽¹⁸⁾ After each round of training, the features corresponding to several weight

coefficients are eliminated, and then the next round of training is performed using a new feature set. The random forest algorithm is a statistical learning theory. The essence is a combined classifier with multiple decisions $\{h(X, \theta_k), k=1, 2, \dots, K\}$, where $\{\theta_k\}$ is a random vector, obeying independent and identical distributions, which determine the form of the decision tree; K is the number of decision trees in the random forest. A simple relative majority voting method is adopted, and the category of the number of decision votes $H(x)$ is used as the category of the final sample, as shown in the equation

$$H(x) = \arg \max_y \sum_k I(h_k(x) = y). \quad (1)$$

2.2 Support vector machine (SVM) regression

In SVM⁽¹⁹⁾ regression, input x is first mapped to the m -dimensional feature space by fixed mapping. The linear model $f(x, w)$ is shown as

$$f(x, w) = \sum_{j=1}^n w_j g_j(x) + b, \quad (2)$$

where $g_j(x), j = 1, 2, \dots, n$, w represents a set of nonlinear transformations, and b is a bias term. Regression estimates can be obtained by minimizing the empirical risk of the training data. SVM regression is shown as

$$\begin{aligned} & \min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*), \\ & \text{subject to} \begin{cases} y_i - f(x_i, w) - b \leq \varepsilon + \xi_i^*, \\ f(x_i, w) + b - y_i \leq \varepsilon + \xi_i^*, \\ \xi_i \xi_i^* \geq 0, \end{cases} \end{aligned} \quad (3)$$

where C is a positive constant (regularization parameter).

2.3 K-means clustering based on contour coefficients

The K-means algorithm⁽²⁰⁾ is a classical clustering method, whose clustering number is a custom parameter. To improve the efficiency of clustering, the silhouette coefficient is introduced to determine the number of clusters.⁽²¹⁾ The silhouette coefficient quantifies the similarity of any object in the dataset to other objects in the clusters and the similarity of the object to objects in other clusters, and combines the two quantized similarities in a certain form to obtain evaluation criteria for the merits of clustering effects. The calculation method for the silhouette coefficient is as shown in the equation

$$silhouette_i = \frac{b_i - a_i}{\max(a_i, b_i)}, \quad (4)$$

where a_i represents the average distance of the i th sample relative to all samples of the cluster in which it is located and b_i represents the minimum average distance of the i th sample and all samples of other clusters. The silhouette coefficient varies between -1 and 1 . When $silhouette_i = 1$, sample i is more dissimilar to samples in other clusters. When $silhouette_i = 0$, the sample i cluster is not obvious. Moreover, when $silhouette_i = -1$, sample i is assigned to an incorrect cluster.

3. Energy Management Strategy of Captive Power Plant

3.1 Captive power plant energy management problem

From a control viewpoint, the energy loss of a captive power plant can generally be divided into controllable and uncontrollable energy losses. Figure 1 shows the energy loss structure. Unit energy consumption is a key indicator of a captive power plant energy management. Using data mining technology to solve the problem of high energy consumption in captive power plants is an important subject.

A captive power plant has more data characteristics, as shown in Table 1, including 68-dimensional variable parameters and key indicator parameters, such as unit energy consumption. The rated power of the power plant is 350 MW; however, owing to the change in load, the power plant does not necessarily operate under the rated operating conditions. To reduce unit energy consumption, the operating parameters are optimized on the basis of different working conditions.

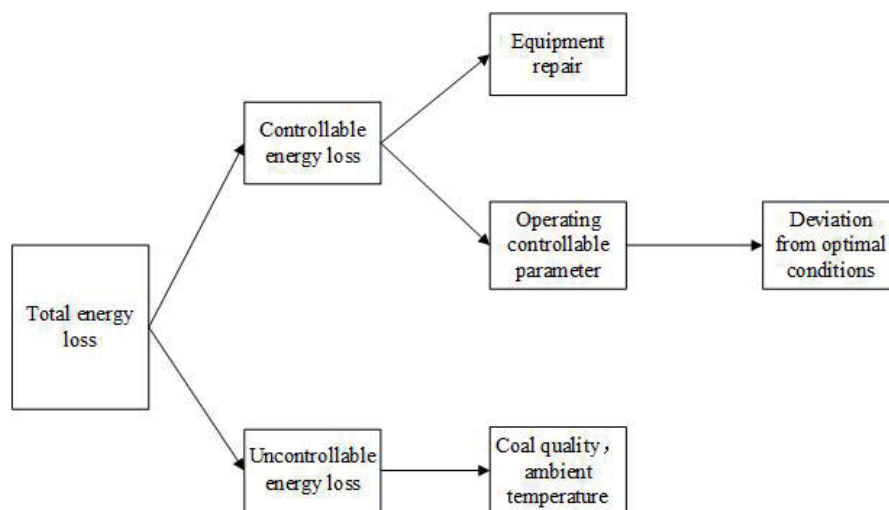


Fig. 1. Energy loss in captive power plants.

Table 1
Data parameter description.

Parameter	Unit	Number
Load	MW	1
Separator temperature	°C	2
Main steam temperature	°C	10
Main steam flow	t/h	1
Main steam fountain temperature	°C	8
Reheater pressure	MPa	4
Reheater temperature	°C	4
Water pressure	MPa	3
Water flow	t/h	2
Water temperature	°C	2
Pressure difference	MPa	4
Total wind	NM ³ /h	1
Furnace temperature	°C	2
Furnace negative	MPa	2
Oxygen from furnace outlet	%	2
Air prebody temperature	°C	12
Received base carbon	%	8
Unit consumption	g/kWh	1

3.2 Modeling framework

As shown in Fig. 2, the overall modeling ideas are as follows. (1) The RF-RFE feature selection method is used to filter data features, and key features are selected. (2) An SVM model is established for unit energy consumption. (3) According to the load interval grouping, the grouped data are selected relative to some optimal historical operating parameters for clustering, and the key parameters are optimized according to the cluster center. (4) The optimized parameters were substituted into the SVM model. Then, the error between the output and the actual unit energy consumption is calculated, and the effectiveness of the cluster optimization method is verified.

3.3 Modeling process

According to the modeling framework proposed in Sect 3.2, this study was implemented in the following four steps:

Step 1: Using the RF-RFE feature selection method, key features are selected; the random forest is a combined model composed of multiple regression decision tree models. The implementation of the RF-RFE algorithm can be summarized as follows: (1) train a random forest from data samples, input features are 68-dimensional parameters, and the output characteristic is unit energy consumption; (2) calculate the importance of features and eliminate irrelevant variables; (3) repeat (1) and (2), until no further variables are maintained.

Step 2: According to the result of step 1, a linear support vector machine model was established for unit energy consumption. The input variables of the model are total coal, total air, water-to-

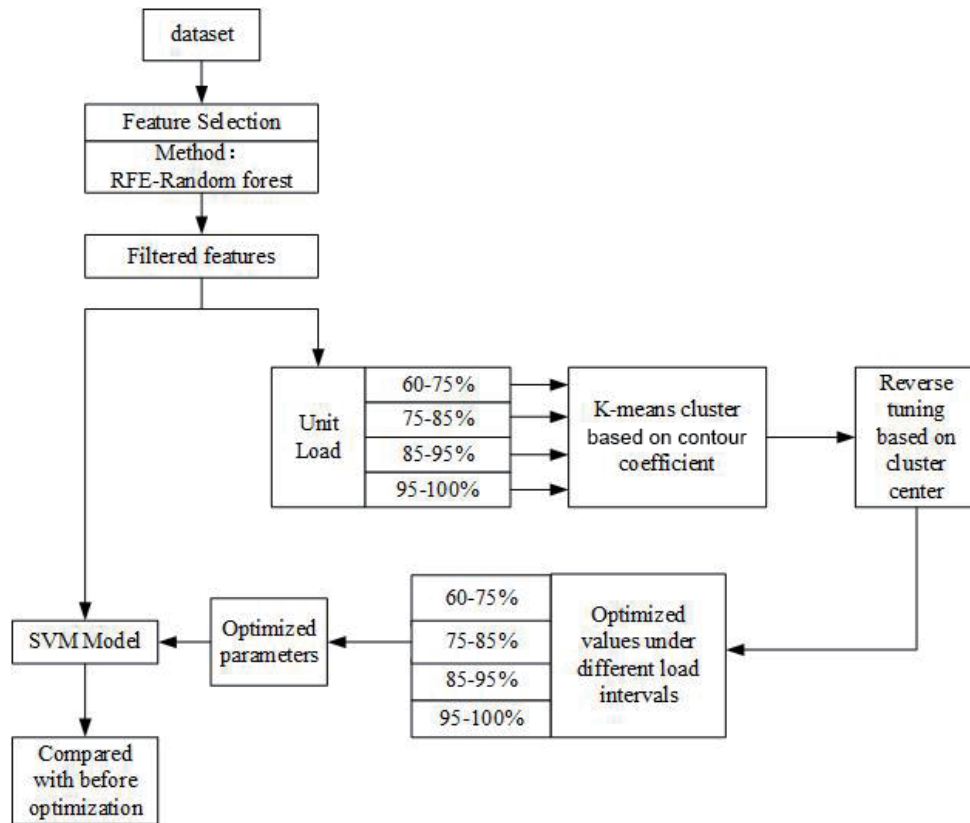


Fig. 2. Overall modeling ideas in this study.

coal ratio, water flow rate, load, output variable which is unit energy consumption, and linear kernel function.

Step 3: According to the result of step 1, historical data sets are grouped on the basis of the load; loads are grouped according to the rated powers of 65–75, 75–85, 85–95, and 95–100%.

The grouped data were selected with some optimal historical data sets for k-means clustering based on the contour coefficient. The data sets were determined using the best number of classifications and the cluster center according to the size of the contour coefficient. On the basis of the cluster center, other key parameters of the corresponding load interval were optimized, and the optimization method is shown as

$$\frac{load[i] - load_{center2}}{load_{center1} - load_{center2}} = \frac{x[i] - x_{center2}}{x_{center1} - x_{center2}}, \quad (7)$$

where $load[i]$ is the load, $load_{center2}$ is one load center constant, $x_{center2}$ is one parameter constant, $load_{center1}$ is another load center constant, $x_{center1}$ is another parameter constant, and $x[i]$ is the variable to be optimized.

The operating parameters can be optimized from Eq. (7) as

$$x[i] = \frac{load[i] - load_{center2}}{load_{center1} - load_{center2}} * (x_{center1} - x_{center2}) + x_{center2}, \quad (8)$$

Step 4: The optimized parameters are substituted into the unit energy consumption linear SVM model, the error between the corresponding model output and the actual unit energy consumption is calculated, and measurements are performed to determine whether unit energy consumption is optimized.

4. Experiment and Analysis

4.1 Experimental data

The experimental data used in the experiment are selected by one month data sampling in the power plant, and a total of 355 sample data sets are obtained. The power plant rated power is 350 MW. Loads are grouped according to the rated powers of 65–75, 75–85, 85–95, and 95–100%. Each group has 57, 31, 40, and 127 sets, as shown in Fig. 3.

4.2 RF-RFE feature selection

The RF-RFE feature selection method is used to select key features from the results shown in Table 2. In Table 2, when the result is true, the parameter is selected. Five characteristics, namely, total coal volume, total air volume, water-to-coal ratio, feed water flow rate, and unit load, are selected as key variables.

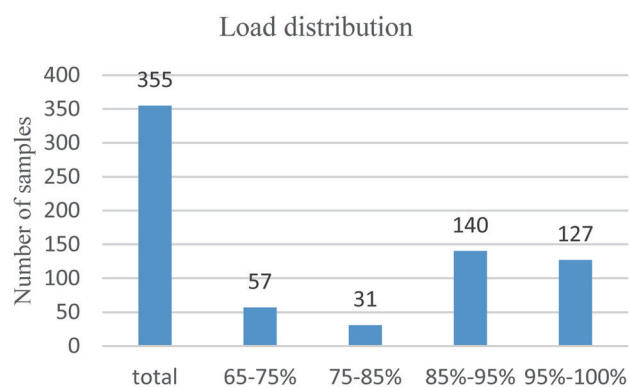


Fig. 3. (Color online) Sample description.

Table 2
RF-RFE feature selection results.

Load	...	Total air	...	Total coal	...	Water-to-coal ratio	Water supply flow
True	False	True	False	True	False	True	True

4.3 Unit energy consumption model based on SVM

The input variables of the model are total coal, total air, water-to-coal ratio, water flow rate, load, output variable, which is unit energy consumption, and linear kernel function.

There are 300 train and 55 test sets. These train and test sets are completely independent of each other. The results are shown in Figs. 4(a) and 4(b), and the mean squared error (MSE) and mean absolute error (MAE) are shown in Table 3.

4.4 Cluster based on contour coefficients

After grouping, the data are selected on the basis of certain historical better working conditions. In accordance with the selected data, we used k-means clustering based on the contour coefficient method. The results are two, two, two, and three classifications, as shown in the cluster center in Table 4.

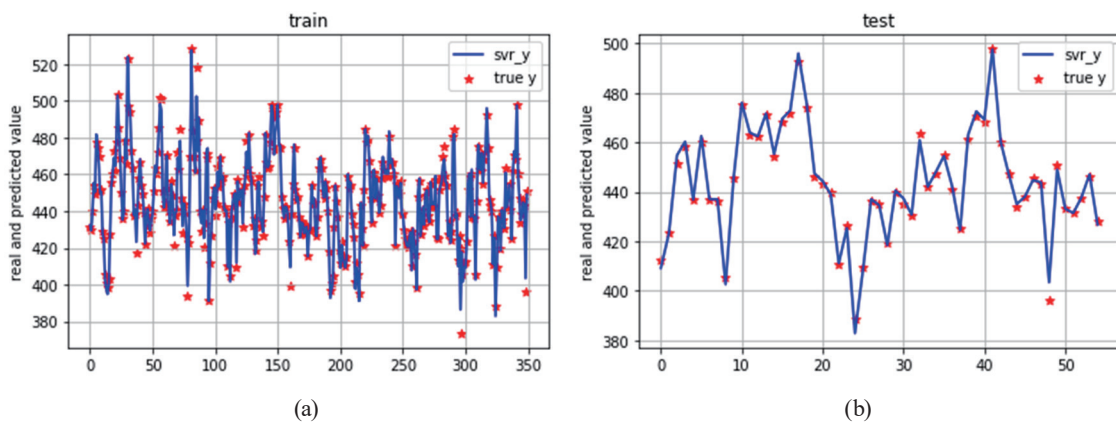


Fig. 4. (Color online) (a) Train and (b) test results.

Table 3
Train and test errors.

Error	Train set	Test set
MSE	3.351	4.628
MAE	1.265	1.540

Table 4
Cluster center.

Parameter	65–75%		75–85%		85–95%		95–100%		
	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2	Class 3
Unit load	239.57	260.56	290.63	276.72	302.67	327.47	341.82	338.98	337.09
Total coal	102.23	113.19	123.58	117.86	130.03	143.17	137.14	135.28	136.39
Total air	647.31	768.50	779.15	767.62	857.84	939.39	936.52	963.30	933.32
Water-to-coal ratio	7.59	7.75	7.96	7.73	7.03	7.58	8.39	8.63	8.23
Water flow	64133.50	83193.90	105424.00	90350.30	117048.40	142203.30	145097.00	149505.00	136905.00

Table 5
Comparison of optimization results.

Interval	Load	Total coal		Total air		Water-to-coal ratio		Water supply		Unit consumption	
		OP-V	ACV	OP-V	ACV	OP-V	ACV	OP-V	ACV	OP-V	ACV
65–75%	239.70	112.53	102.30	683.90	648.11	6.94	7.59	64212.03	64257.80	469.44	443.39
	240.66	108.10	102.80	724.90	653.61	7.22	7.61	65173.34	65123.30	449.17	430.34
	245.54	116.67	105.35	647.50	681.78	7.30	7.64	77044.67	69554.40	475.15	450.30
	290.08	145.57	123.35	824.41	778.69	7.04	7.95	114810.10	104822.00	501.83	435.25
75–85%	281.23	125.38	119.71	805.69	771.35	7.41	7.81	92623.90	95229.60	445.83	426.89
	278.06	113.90	108.41	700.13	738.73	8.02	7.75	93325.81	91797.90	419.62	407.47
	323.21	152.12	140.91	984.77	925.39	7.56	7.49	144534.20	137887.00	470.64	441.31
85–95%	319.28	152.53	138.83	979.21	912.48	7.38	7.44	137759.30	133904.00	477.75	445.61
	325.61	147.06	142.18	967.82	933.29	7.77	7.54	142886.30	140324.00	451.64	421.13
	340.60	136.55	134.33	955.72	989.04	8.41	8.97	144458.00	160330.00	420.91	397.98
95–100%	339.15	136.65	135.18	991.40	966.03	8.54	8.67	150180.10	150654.00	402.93	389.98
	350.60	161.41	156.19	1020.00	1010.18	7.19	7.37	146411.10	145938.10	460.36	445.61

4.5 Optimization and analysis

The load is an uncontrollable variable, depending on the external requirements of the plant. Total coal, total air, water-to-coal ratio, and water flow rate are operating parameters of the power plant. By adjusting the corresponding parameters, the unit energy consumption can be reduced and the energy loss of the power plant can be minimized. According to the cluster center, the load variable was used as a reference standard, the manipulated variable parameters were tuned, and the optimized parameters were input into the support vector machine model. In this study, three sets of samples with different load intervals were selected, and a total of 12 sets of samples were tested. The optimization results were compared with the actual running results. The comparison results are shown in Table 5.

In Table 5, ACV represents the actual operating parameter and OP-V represents the parameter after optimization. According to the unit energy consumption results in Table 5, other key parameters were optimized according to the load requirements of different intervals. The optimized results show that the unit energy consumption can be reduced by 3–5%. For different self-supplied power plant load requirements, the goal of saving unit energy consumption can be achieved.

5. Conclusions

In this study, the RF-RFE feature selection method was used for feature selection, and some data with historical parameters were selected according to different load intervals for clustering. According to the cluster center, tunable parameters were optimized, and optimization results were input into the SVM model for energy consumption. Comparison results show that the optimized parameters can reduce the energy consumption of the power plant and conserve energy, which has a certain engineering guiding significance.

References

- 1 G. Li and P. F. Niu: *Soft Comput.* **20** (2016) 1. <https://doi.org/10.1007/s00500-014-1486-3>
- 2 J. Z. Chu, S. S. Shieh, S. S. Jang, C. I. Chien, H. P. Wan, and H. H. Ko: *Fuel* **82** (2003) 693. [https://doi.org/10.1016/S0016-2361\(02\)00338-1](https://doi.org/10.1016/S0016-2361(02)00338-1)
- 3 P. Ilamathi, V. Selladurai, and K. Balamurugan: *J. Energy Res. Technol.* **135** (2013) 032201. <https://doi.org/10.1115/1.4023328>
- 4 J. Smrekar, M. Assadi, M. Fast, I. Kuštrín, and S. De: *Energy* **34** (2009) 144. <https://doi.org/10.1016/j.energy.2008.10.010>
- 5 G. Li, P. Niu, H. Wang, and Y. Liu.: *Neural Networks* **51** (2014) 57. <https://doi.org/10.1016/j.neunet.2013.12.006>
- 6 B. S. Kim, T. Y. Kim, T. C. Park, and Y. K. Yeo: *Korean J. Chem. Eng.* **35** (2018) 1779. <https://doi.org/10.1007/s11814-018-0087-8>
- 7 Y. Cheng, Y. X. Huang, P. Bo, and Z. Weidong: *Eng. Appl. Artif. Intell.* **74** (2018) 303. <https://doi.org/10.1016/j.engappai.2018.07.003>
- 8 A. A. M. Rahat, W. Chunlin, R. M. Everson, and J. E. Fieldsend: *Appl. Energy* **229** (2018) 446. <https://doi.org/10.1016/j.apenergy.2018.07.101>
- 9 C. Wang, Y. Liu, S. Zheng, and A. Jiang: *Energy* **153** (2018) 153. <https://doi.org/10.1016/j.energy.2018.01.003>
- 10 F. Wang, S. X. Ma, H. Wang, Y. D. Li, and J. J. Zhang: *Control Eng. Prac.* **80** (2018) 26. <https://doi.org/10.1016/j.conengprac.2018.08.003>
- 11 H. Zhao, J. Shen, Y. Li, and J. Bentsman: *Control Eng. Prac.* **58** (2017) 127. <https://doi.org/10.1016/j.conengprac.2016.10.005>
- 12 M. Li, F. Hou, and Q. Zhou: *Smart Grid Technologies Asia (IEEE, 2016)* 1–5. <https://doi.org/10.1109/ISGT-Asia.2015.7387118>
- 13 P. Tan, J. Xia, C. Zhang, and Q. Fang: *Energy* **94** (2016) 672. <https://doi.org/10.1016/j.energy.2015.11.020>
- 14 S. H. Baek, H. Y. Park, and S. H. Ko: *Fuel* **128** (2014) 62. <https://doi.org/10.1016/j.fuel.2014.02.043>
- 15 B. H. Lee and E. G. Eddings: *Energy Fuels* (2012) 121031112740002. <https://doi.org/10.1021/ef300562t>
- 16 S. Helle, A. Gordon, G. Alfaro, X. Garcia, and C. Ulloa: *Fuel Process. Technol.* **80** (2003) 209. [https://doi.org/10.1016/S0378-3820\(02\)00245-X](https://doi.org/10.1016/S0378-3820(02)00245-X)
- 17 H. G. Liu: *Huadian Technol.* **31** (2009) 34. <http://www.hdpower.net/EN/Y2009/V31/I9/34>
- 18 C. Qi, M. Zhaopeng, L. Xinyi, J. Qianguo, and S. Ran: *Genes* **9** (2018) 301. <https://doi.org/10.3390/genes9060301>
- 19 H. Yang, L. Chan, and I. King: *Intelligent Data Engineering and Automated Learning — IDEAL* **2412** (2002) 391. https://doi.org/10.1007/3-540-45675-9_58
- 20 J. A. Hartigan and M. A. Wong: *J. R. Stat. Soc.* **28** (1979) 100. <https://doi.org/10.2307/2346830>
- 21 S. X. Xia, W. C. Li, Y. Zhou, L. Zhang, and Q. Niu: *J. Southeast Univ.* **23** (2007) 435. <https://doi.org/10.3969/j.issn.1003-7985.2007.03.027>

About the Authors



Juan Liu received her B.S. degree in automation from North China University of Water Resources and Electric Power, Zhengzhou, China, in 2017. Since 2017, she has been studying at Shanghai Maritime University. Her current interests include industrial data analysis and machine learning.



Xixia Huang received her B.E. degree from Harbin Engineering University, Harbin, China, in 1997, her M.E. degree from Harbin Institute of Technology, Harbin, in 1999, and her Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2008. She is currently an associate professor at Shanghai Maritime University, Shanghai. Her research interests include intelligent manufacturing and industrial data analysis.



Xiaoli Liu received her M.S. degree in medical test from Guangzhou University of Chinese Medicine, Guangzhou, China, in 2017. Her current interests include industrial data analysis and machine learning.



Zhiliang Han received her B.E. degree from Henan Polytechnic University, China, in 2016. Since 2017, she has been studying at Shanghai Maritime University for her M.E. degree. Her research interests include computer vision, image processing, pattern recognition, and machine learning