Energy Consumption Optimization of the Aluminum Industrial Production Based on Pattern Recognition

Xiaofang Lou
Department of Automatic Control, College of Mechatronics and Automation
National University of Defense Technology, Changsha 410072, China
E-mail: jiangxi000000@sohu.com

Fengxing Zou
Department of Automatic Control, College of Mechatronics and Automation
National University of Defense Technology, Changsha 410072, China
Tel: 86-731-84574370

The research is financed by natural science fundation of China (Grant: 60634042); and supported by the graduate innovation fund of the National University of Defense Technology (Grant: S090304)

Abstract

The basic energy consumption situation in 1 ton of aluminum production is addressed in this paper for the current aluminum industry from mining to the aluminum processing. In order to save energy of the aluminum industrial production, the K-means algorithm and threshold cluster analysis algorithm from the pattern recognition are proposed, and the classification of the energy consumption data of the aluminum industrial production is completed by using C++. The two algorithms are compared in terms of minimum total energy consumption and the actual production requirements. A class is found with more data number and smaller total energy consumption data, and such class heart is used as an input to save the aluminum industrial production. Further-more the feasibility and the effectiveness of the algorithms are verified.

Keywords: Aluminum industry, Energy consumption, Pattern recognition, K-means algorithm, Threshold method, Cluster analysis

1. Introduction

Energy consumption of the aluminum industrial processes has always been a problem affecting people, which have been continually studied and explored. Despite many successes, there are still many gaps to do. Pattern recognition theory and methods have been a widespread attention in many science and technology fields. It has promoted the development of artificial intelligence system, and has expanded the areas of computer applications. For years, pattern recognition research has already made great achievements. This paper is aimed to classify the energy consumption data of the aluminum industrial production, extracting the implicit rules, and guiding the practical production, so as to save energy.

This paper introduces the relevant knowledge of pattern recognition and cluster analysis, extracts the energy consumption situation of the aluminum industrial production, and proposes two cluster analysis algorithms for a large number of energy consumption data in the aluminum industrial production. The energy consumption data is classified by using C++, and an energy saving way is found.

2. Pattern recognition and cluster analysis

2.1 Pattern recognition

Pattern recognition is a process that describes, identifies, classifies and explains the thing or phenomenon, through analyzing all the information which indicates the thing or phenomenon. Its main application fields include such subjects as image analysis, voice recognition, sound classification, communication, computer aided diagnosis, data mining and so on. According to the property of the problem and the method solving the problem, pattern recognition can be divided into classification with supervision and classification without supervision. The major difference between them is that whether the categories of the experimental specimens are known in advance. Generally speaking, the classification with supervision is often necessary to provide a large number of samples whose categories are known, but in actual problems, it often has certain difficulties. It is often needed to research the classification without supervision(Li Jing-jiao translate, 2006). There are two basic pattern recognition methods, statistical pattern recognition method and structural pattern recognition method. The statistical pattern recognition method is based on the probability density function for the distribution of the pattern set in the character space, then do some research on the general character, cluster analysis is a kind of classification without supervision.

2.2 Cluster analysis (Yin Bo, 2008)

The process of dividing the physical or abstract objects into a group of classes or domains is called clustering. The class generated by clustering is a collection of objects, which are similar to the objects of the same class, and different to the objects of the other class. So in many applications, the data objects of the same class can be treated as a whole.

The mathematical description of clustering process is as follows(Yang Xiao-bing, 2005):

Define the sample set to be studied as$X$, and a non-empty subset of X class$C$, that is, $C \in X$ and$C \neq \phi$.

The ultimate goal of clustering is to divide the sample set X into k partitions$C_m(m = 1, ..., k)$, there may be some objects not belonging to any partition, which is defined to belong to the noiseC_n. The union of all the partitions and noise is the sample set$X$, and there is no intersection between these partitions, that is:

1) $C_1 \cup C_2 \cup ... \cup C_k \cup C_n = X$

2) $C_i \cap C_j = \phi$ (For any$i \neq j$)

The partitions C_m are clustering. The second condition says that each sample of the sample set X belongs to one class.

3. Cluster analysis algorithm

Cluster analysis algorithm depends on the data type, the purpose and the application of the cluster analysis. This paper proposes two cluster analysis algorithms as follows.

3.1 The threshold cluster analysis algorithm(Sun Ji-xiang, 2001)

1) The terms and conventions

Set up the pattern set to be classified as$\{x_1, x_2, ..., x_N\}$, select the within-class distance threshold *T*.

2) Algorithm idea

Calculate the distance from the pattern feature vector to the cluster center, and compare it with the threshold *T* , then decide which class it belongs to or if a new class center is needed to establish. Using Euclidean Distance in the algorithm:

Set$x = (x_1, x_2, ..., x_n)'$, $y = (y_1, y_2, ..., y_n)'$. Euclidean Distance:

$$d(x, y) = \|x - y\| = \left[\sum_{i=1}^{n} (x_i - y_i)^2\right]^{1/2}$$

3) The principle and steps of the algorithm

(1)Take any pattern feature vector as the first cluster center. For example, command $z_1 = x_1$ is the center of classw_1.

(2)Calculate the distance d_{21} from the next pattern feature vector$x_2$toz_1. If$d_{21} > T$, then establish a new class w_2, with its center being$z_2 = x_2$; if$d_{21} \leqslant T$, then$x_2 \in w_1$.

(3)Assuming cluster centers have been$z_1, z_2, ..., z_k$, calculate the distanced_{ij} from the unsettled pattern feature vector x_i to each cluster center$z_j(j = 1, 2, ..., k)$. If$d_{ij} > T$, then set x_i as the center of a new classw_{k+1}, $z_{k+1} = x_i$; Otherwise, if the

$$d_{il} = \min_j [d_{ij}]$$

then$x_i \in w_l$. Check whether all of the patterns are finished. If finished, then come to the end; otherwise return to (3).

3.2 K-means algorithm with error limit(Li Jie-gu, & Cai Guo-lian, 1986)

K-means algorithm makes the sum of the squared distance from all the samples in the domain to the cluster center minimum, which is based on the criterion function of the sum of the squared error. K-means algorithm consists of the following steps:

1) Define K value and error limit, arbitrarily choose K initial cluster centers$z_1(1), z_2(1), ..., z_K(1)$. Usually select the first K samples from the given sample set as the initial cluster centers.

2) The kth iteration. If$\|x - z_j(k)\| < \|x - z_i(k)\|$,$i = 1, 2, ..., K, i \neq j$, then $x \in f_j(k)$ the cluster center, it is the sample set of $z_j(k)$. Repeat this step until each sample x is distributed to the K cluster domains.

3) Results from the second step, calculate the new cluster center

$$z_j(k+1) = \frac{1}{N_j} \sum_{x \in f_j(k)} x, j = 1, 2, ..., K$$

Which leads to the criterion function of the minimum sum of the squared error.

$$J_j = \sum_{x \in f_j(k)} \|x - z_j(k+1)\|^2, j = 1, 2, ..., K$$

When the gap between the calculated new cluster center and the original cluster center is greater than the pre-defined error limit, then update the cluster center, otherwise maintain the original cluster center.

4) If$z_j(k+1) = z_j(k) j = 1, 2, ..., K$, algorithm converges, and the program ends. Otherwise, go to the second step. Count the data number of each class when classifying.

4. Cluster analysis algorithms applied to the aluminum industrial production

Aluminum industrial production system includes mining plant, alumina plant, electrolytic aluminum plant and aluminum processing plant. Each procedure consumes a lot of energy.

Mining plant is mainly to transport mine. The current alumina production use bauxite, nepheline, and alunite[6]. For example, alunite has the alumina content of 37%(Wang Yan-ming, Yan Ding-ou, Yang Chong-yu, & Gao Shou-ye translate, 1987). 1 ton of aluminum production needs approximately 5.1 tones of alunite. According to the present fossil fuel consumption of transportation, transporting so much alunite about 1 km consumes fuel at cost 0.25 Yuan.

Alumina plant: Comprehensive energy consumption of 1 ton of alumina production is shown in Table 1(Wu Jin-wei, 2003). According to the chemical expression of alumina, 1 ton of aluminum production probably needs 2 tons of alumina, we can get the energy consumption of standard coal in 1 ton of aluminum production, as shown in Table 2. The alumina plant in the production of 1 ton of aluminum requires energy is shown in Table 3.

Electrolytic aluminum plant is mainly carried out through the electrolytic cell. In the electrolytic cell, the expression of the energy consumption (W) is (Dong Shi-yi, Liu Yong-qiang, & Li Wei-bo, 2007):

$$W=2980V/\eta \text{ (kWh/t-Al)}$$

Where V is the average voltage of the electrolytic cell, η is the current efficiency. According to (Yang Yun-bo, & Dong Chun-ming, 2009), the average current efficiency of Chinese advanced aluminum plant is 94.9%. Reference (Yang Ling, 2008) points out that the current average voltage of electrolytic aluminum plant is 4.22V. Therefore, through (1), the energy consumption in the production of 1 ton of aluminum is calculated as 13251.4kWh. The coefficient of standard coal for electric is 0.407 according to Table 1. Thus the energy consumption of the electrolytic aluminum plant in the production of 1 ton of aluminum is calculated as 5393.3kg standard coal.

In the aluminum processing plant, the main production processes include casting, forging, oxidation of color, etc (Xiao Cui-ping, 2005). The energy consumption of Chinese major aluminum processing plant is given in Table 4(Xiao Cui-ping, 2005).

For the aluminum processing plant that has only three processes and the products of each process(Xiao Cui-ping, 2005), energy consumption of the aluminum processing plant can be simplified as the energy consumption in the production of ingot, aluminum alloy plate and oxidation coloring material.

As the foregoing analysis, the situation of energy consumption in the production of 1 ton of aluminum, including mining plant, alumina plant, electrolytic aluminum plant and aluminum processing plant, is listed in Table 5. There are 10 factors (component 1-10) affecting the total energy consumption.

A batch of stochastic data is generated by Matlab according to Table 5. It is based on certain distributions in Table 6. *N* means normal distribution, *U* means continuous uniform distribution. Finally each data vector has 11 components. See the first 10 components in Table 5. The last component is the sum of the first 10 components, the total energy consumption.

5. Simulation results and analysis

The classification of a lot of energy consumption data of the aluminum industrial production is completed by using the two cluster analysis algorithms mentioned in subsection 3. The threshold *T* is determined through explore method, defines different thresholds, compare and analyze the different results. When *T* is larger than 550, get too many classes, when *T*=550,600,700 and 800, the classification results respectively are shown in Table 7, Table 8, Table 9 and Table 10. E means the total energy consumption of the class, Num means the data number of the class, and Error represents the error limit in subsection 3.

For K-means algorithm define different *K* values and error limits, when the error limits are defined as 0.001, 0.1 and 100, the classification results for all the different *K* values are the same, so for all the different *K* values respectively define three error limits of 0.001,10 and 100, when *K*= 5 and *K*= 20, the classification results are shown in Table 11 and Table 12, when *K*= 10 and *K*= 15, the classification results of error limit of 0.001 and 10 are the same, so the classification results are shown in Table 13 and Table 14 only have two cases. If the data number of the class is large, it shows that the class in reality is relatively easy to implement. In order to achieve energy saving of the aluminum industrial production, we need to find the class with minimum total energy consumption.

When *K*=20, for different error limits the largest data number are respectively 75, 69, 78, the maximum energy consumption are respectively 9.998, 9.998, 9.981, the minimum energy consumption are respectively 7.283, 7.283, 7.401; While the threshold cluster analysis algorithm when *T*=550, in the 18 classes, the largest data number is 159, the maximum energy consumption is 10.156, the minimum energy consumption is 7.167. When *K*=15, for the two cases the largest data number are both 101, the maximum energy consumption are respectively 9.848, 9.931, the minimum energy consumption are respectively 7.572, 7.600; While the threshold cluster analysis algorithm when *T*=600, in the 14 classes, the largest data number reach 218, the maximum energy consumption is 10.110, the minimum energy consumption is 6.990. When *K*=10, for the two cases the largest data number are respectively 128, 147, the maximum energy consumption are respectively 9.800, 9.699, the minimum energy consumption are respectively 7.610, 7.777; While the threshold cluster analysis algorithm when *T*=700, in the 11 classes, the largest data number reach 305, the maximum energy consumption is 10.074, the minimum energy consumption is 6.990. When *K*=5, for the three error limits the largest data number are respectively 287, 285, 267, the maximum energy consumption are respectively 9.492, 9.499, 9.409, the minimum energy consumption are respectively 7.785, 7.803, 7.865; While the threshold cluster analysis algorithm when *T*=800, in the 5 classes, the largest data number reach 404, the maximum energy consumption is 9.750, the minimum energy consumption is 7.473. For K-means algorithm, the more the classes, the stronger the regularity, and the classification results of different error limits are nearly the same, so the superiority is not obvious; But for the threshold cluster analysis algorithm for different thresholds the data number are far more different from each other, and the total energy consumption has a large span, so the superiority is rather obvious. As the cluster centers of the threshold cluster analysis algorithm are all from the original data, and once the cluster centers are determined, they will not change, when *T*=600 and *T*=700, we get the minimum total energy consumption 6.990, and the data number is just 1, so it's the minimum total energy consumption of the original data, therefore this algorithm can also find the minimum total energy consumption of the original data.

Contrast, and analyze the classification results, considering the actual production, in order to save energy in the aluminum industrial production, this paper chose the third class of the result when *T*=800, use the first 10 components of the cluster center (0.053, 0.211, 0.612, 0.227, 0.022, 0.006, 4.885, 0.396, 0.903, 0.758, 8.073) as input, which can reach the purpose of energy saving while fulfilling the task of the aluminum industrial production.

6. Conclusion

In this paper, energy saving of the aluminum industrial production is discussed, that is to say, the basic energy consumption situation in 1 ton of aluminum production, for the current aluminum industry from mining to the aluminum processing. The K-means algorithm and the threshold cluster analysis algorithm are proposed, and the classification of a lot of energy consumption data of the aluminum industrial production is completed by using C++. After analyzing the different classification results, it shows that for the K-means algorithm, the more the classes, the stronger the regularity, so the superiority is not obvious, but for the threshold cluster analysis algorithm, the superiority is rather obvious. When T = 800, the third class of the classification result is the best, and the total energy consumption of the class is only 8.073, with the data number of the class being 210. Using the first 10 components of the cluster center (0.053, 0.211, 0.612, 0.227, 0.022, 0.006, 4.885, 0.396, 0.903, 0.758, 8.073) as input, energy saving of the aluminum industrial production is greatly achieved. The feasibility and effectiveness of the algorithm are verified through simulations.

References

Dong Shi-yi, Liu Yong-qiang, & Li Wei-bo. (2007). Discussion on ways of reducing voltage in pre-baked anode aluminum electrolysis cell. *Technology and equipment*, 6, 23-24.

Li Jie-gu, & Cai Guo-lian. (1986). Computer pattern recognition techniques. Shanghai Jiao-tong University Press. 54-55.

Li Jing-jiao translate. (2006). Pattern Recognition. Electron Industry Press.

Sun Ji-xiang. (2001). Modern Pattern Recognition. Changsha: National University of Defense Technology Press. 26.

Wang Yan-ming, Yan Ding-ou, Yang Chong-yu, & Gao Shou-ye translate. (1987). Alumina Production Process and Equipment. Beijing: Metallurgical Industry Press. 3,5.

Wu Jin-wei. (2003). Energy saving measures and effects of alumina reconstruction in Guizhou Aluminum Works. *Energy Saving of Non-ferrous Metallurgy*, 20, 1-4.

Xiao Cui-ping. (2005). On energy saving way for aluminum processing plant. *Ke-yuan forum*, 6, 39-41.

Yang Ling. (2008). The approach of energy saving and consumption reduction on 200kA large aluminum reduction cell. *Energy for Metallurgy industry*, 27, 26-29.

Yang Xiao-bing. (2005). Research of several key technologies of cluster analysis. Hangzhou: Zhejiang University. 23-28.

Yang Yun-bo, & Dong Chun-ming. (2009). Preliminary investigation and analysis of energy efficiency in China's electrolytic aluminum industry. *Energy Saving of Non-ferrous Metallurgy*, 5, 5-7.

Yin Bo. (2008). Cluster analysis and its application research in the data mining analysis of mobile communications business. Changsha: Hunan University. 9-10.

Table 1. Comprehensive energy consumption of 1 ton of alumina production

No	Energy	Unit	Quantity	Coefficient of standard coal	Standard coal (kg)
1	Coke	Kg	27.0	0.971	26.2
2	Gas	Nm3	596.0	0.188	106.1
3	High-pressure Steam	Kg	2100.0	0.096	201.6
4	Low-pressure Steam	Kg	1140.0	0.094	107.2
5	Electric	KWh	280.0	0.407	114.0
6	New Water	m3	10.0	0.200	2.0
7	Recycling Water	m3	90.0	0.100	9.0
8	Compressed Air	m3	80.0	0.040	3.2
	Total				569.3

Table 2. Comprehensive energy consumption of standard coal in 1 ton of aluminum production

Energy	Coke	Gas	High-pressure steam	Low-pressure steam	Electric	New water	Recycling water	Compressed air	Total
Standard Coal(kg)	52.4	212.2	403.2	214.4	228	4.0	18.0	6.4	1138.6

Table 3. Energy consumption of alumina plant in the production of 1 ton of aluminum

Energy	Coke	Gas	Steam	Electric	Water	Air
Standard Coal (kg)	52.4	212.2	617.6	228.0	22.0	6.4

Table 4. Energy consumption of aluminum product (unit: kg standard coal / t)

Product name	Ingot	Casting plate	Aluminum alloy plate	Aluminum foil	Oxidation coloring material
Energy Range of Process	258~676	156~370	595~1309	282~1617	497~1069

Table 5. Energy consumption situation of aluminum industry in the production of 1 ton of aluminum (standard coal (kg))

Mining Plant	A small amount of energy may be negligible					
Alumina Plant	Coke(component 1)	Gas(component 2)	Steam(component 3)	Electric(component 4)	Water(component 5)	Air(component 6)
	52.4	212.2	617.6	228.0	22.0	6.4
Electrolytic Aluminum Plant	Electric (component 7)					
	5393.3					
Aluminum Processing Plant	Casting(component 8)		Forging (component 9)		Oxidation of color (component 10)	
	258~676		595~1309		497~1069	

Table 6. Stochastic data

Distribution Function	Coke	Gas	Steam	Electric	Water
	N (52.4,5)	N (212.2,20)	N (617.6,60)	N (228,20)	N (22,2)
	Air	Electric	Casting	Forging	Oxidation of color
	N (6.4,0.6)	N (5393.3,500)	U (258,676)	U (595,1309)	U (497,1069)
Data Number	1000				

Table 7. When $T = 550$, the statistics of classification (unit: 10^3kg standard coal)

Class	1	2	3	4	5	6	7	8	9
E	8.605	9.314	7.780	8.387	9.685	8.932	8.120	9.199	7.412
Num	159	108	74	78	41	88	85	63	3
Class	10	11	12	13	14	15	16	17	18
E	7.167	8.775	8.314	8.117	10.156	8.975	9.684	9.874	7.397
Num	9	134	63	27	8	43	7	3	7

Table 8. When T = 600, the statistics of classification (unit: 10^3kg standard coal)

Class	1	2	3	4	5	6	7
E	8.574	9.270	7.866	9.715	8.153	9.190	7.412
Num	218	126	95	45	125	72	3
Class	8	9	10	11	12	13	14
E	7.334	8.775	8.980	8.400	10.110	9.676	6.990
Num	19	170	65	48	8	5	1

Table 9. When T = 700, the statistics of classification (unit: 10^3kg standard coal)

Class	1	2	3	4	5	6	7	8	9	10	11
E	8.519	9.261	8.002	9.505	7.444	8.827	10.074	8.947	9.693	6.990	8.422
Num	305	142	175	65	33	169	15	81	7	1	7

Table 10. When T = 800, the statistics of classification (unit: 10^3kg standard coal)

Class	1	2	3	4	5
E	8.629	9.138	8.073	9.750	7.473
Num	404	286	210	60	40

Table 11. When K= 5, the statistics of classification (unit: 10^3kg standard coal)

Error	Class	1		2		3		4		5	
		E	Num	E	Num	E	Num	E	Num	E	Num
0.001		8.847	203	8.886	203	9.492	170	7.785	137	8.357	287
10		8.909	285	8.689	196	9.499	164	7.803	143	8.315	212
100		8.897	267	8.599	183	9.409	210	7.865	169	8.332	171

Table 12. When K= 10, the statistics of classification (unit: 10^3kg standard coal)

Error	Class	1	2	3	4	5	6	7	8	9	10
0.001	E	8.757	8.521	9.092	7.610	8.212	9.057	8.195	9.436	9.800	8.690
	Num	117	113	86	78	128	122	101	87	46	122
100	E	8.770	8.560	9.353	7.777	8.280	9.056	8.184	9.699	9.591	8.724
	Num	116	138	85	120	92	147	97	27	58	120

Table 13. When K= 15, the statistics of classification (unit: 10^3kg standard coal)

Error	Class	1	2	3	4	5	6	7	8
0.001	E	8.705	8.525	9.366	8.165	8.329	9.098	7.971	9.664
	Num	91	101	76	68	66	70	75	37
100	E	8.780	8.518	9.369	8.126	8.250	9.097	8.111	9.698
	Num	76	95	69	69	74	101	57	29
Error	Class	9	10	11	12	13	14	15	
0.001	E	9.144	8.685	8.961	9.848	8.445	8.879	7.572	
	Num	56	62	74	27	69	65	63	
100	E	9.362	8.686	8.980	9.931	8.495	8.807	7.600	
	Num	48	81	59	20	73	73	76	

Table 14. When K= 20, the statistics of classification (unit: 10^3kg standard coal)

Error	Class	1	2	3	4	5	6	7	8	9	10
0.001	E	8.756	8.404	9.409	8.102	8.308	9.057	8.410	9.731	9.600	8.680
	Num	56	75	56	57	48	62	56	22	25	59
10	E	8.757	8.390	9.408	8.148	8.308	9.054	8.369	9.719	9.600	8.651
	Num	53	69	57	54	49	65	51	23	25	59
100	E	8.799	8.492	9.399	8.159	8.318	9.097	8.217	9.678	9.598	8.634
	Num	49	78	58	59	58	66	38	28	23	60
Error	Class	11	12	13	14	15	16	17	18	19	20
0.001	E	8.999	9.998	8.500	8.715	7.283	7.983	8.791	9.234	7.728	9.055
	Num	52	16	42	68	19	55	55	56	57	54
10	E	9.004	9.998	8.574	8.708	7.283	7.986	8.745	9.200	7.730	9.051
	Num	51	16	43	67	19	61	66	59	58	55
100	E	8.975	9.981	8.514	8.703	7.401	7.998	8.767	9.183	7.744	9.009
	Num	49	17	59	52	28	52	69	58	50	49