Data analysis of blast furnace gas center based on STING grid clustering

Ming-xin LI¹, Xiu-qin WANG²

¹School of Science, University of Science and Technology of Inner Mongolia, Baotou 014010, Inner Mongolia.
²chofn (Tianjin) Intellectual Property Agency Co., Ltd, Tianjin, 300000.

Ming-xin Li, Doctor, lecturer; E-mail: mingxin1581@163.com;
Xiu-qin WANG, E-mail: 1248204503@qq.com;

Abstract: Distribution of gas flow is related to the temperature distribution in the furnace, furnace condition running, and utilization of the gas, ultimately affecting the blast furnace smelting. The research results show that the gas utilization and temperature field center shift gas share were positively correlated, and negatively correlated with the excursion of the larger and smaller gas share. After the processing of the STING grid, there are five categories. The comparison shows that the gas utilization rate is mainly determined by the non offset gas occupation rate of the temperature field center.

1. Introduction.
The distribution of gas flow is related to the distribution of the temperature inside the furnace, the smooth operation of the furnace condition and the status of utilization for the gas, finally it will impact the smelting index of blast furnace. The operation of blast furnace is conducted with the focus on the reasonable and suitable distribution of gas flow. On the other hand, the distribution of gas flow is also the important accordance for the operators of blast furnace to judge for the furnace condition[1]. Our purpose is to establish the clustering model, to explore the relationship between the distribution mode of gas flow and the utilization ratio of gas for the realization of effective online control for the furnace condition of blast furnace.

2. Data process

2.1 Data preprocess
This article collects 659 groups of data from the blast furnace of Baotou Steel Corp with 4 indexes which are the gas shares with no position shift, the gas shares with little position shift, the gas shares with larger position shift and the gas shares at the center of the temperature field, it marks each group of data with number 1, 2, 3, ..., 659 and forms the initial sample group. We conduct the data process with 3 σ criterion(Pauta criterion) in the statistic discrimination. Suppose this data $X = (X_1, X_2, X_3, X_4)$, the deviation of a certain sample data $x_i$ $(i = 1, 2, ..., n)$ and the standard deviation can be calculated in accordance with the Bayes Formula:
Then it can be considered that is abnormal data which shall be eliminated. Through the process for the abnormal value of data, we eliminate the data that does not meet the standard, the remained 639 groups of data will be used as the new initial sample group.

2.2 The analysis on the correlation of data

Generally the correlation between the variables can be indicated with the coefficient of correlation, the most outstanding feature of the method for the analysis on the coefficient of correlation is that it does not make a difference for the primary and secondary of the variables, it puts all the variables at the same position.

Through correlation analysis, it is concluded that X1 and X4, X2 and X3 appears as positive correlation, X4 and X2, X3 appears as negative correlation. Therefore, we can get the conclusion: X4 (gas shares) was impacted with larger positive correlation from X1 (no offset) of gas shares, it was impacted with little negative correlation from X2 (smaller offset), it was impacted with larger negative correlation from X3 (larger offset).

It is known that $X_1 + X_2 + X_3 = 1$, we only conduct the research on $X_1$ and $X_2$ for the sample data with the clustering algorithm in this article.

3. Clustering model

3.1 Clustering analysis

The content of clustering analysis\(^{[1]}\) is very rich, It can be divided into partitioning method, hierarchical method, Density-based clustering, mesh clustering, model clustering in accordance with the method of clustering. This article mainly applies with the partitioning method, the density-based clustering and the mesh clustering. This article conducts the research on the algorithm of STING\(^{[2]}\), it conducts the analysis of comparison for the two methods of K-means\(^{[3]}\) clustering and dbscan\(^{[4]}\) density-based clustering.

3.2 The method of STING mesh clustering

The algorithm of STING is a multi-resolution clustering based on mesh, it divides the space region into matrix unit of different grades in accordance with the resolution of different grades and forms a level construction. The mother unit of higher grade is received through the sub-unit of multiple low grades, the unit of lower grade continues to contain the sub-unit that is lower than it for the grade. As it is shown in fig 1:

![Hierarchical distribution structure graph](image)

The statistical information for the attributes of each mesh unit has been calculated and stored in advance. These statistical variables can be checked at any time for the convenience of the querier to use. The statistical parameters of the mother unit are received through the information accumulation for the sub units, these parameters include: the parameter-count that is not related to attributes, the
parameter -mean that is related to attributes, the standard deviation (stdev), the minimum value (min), the maximum value (max) and the distribution type that is followed by the value of attributes in this unit, such as the consistent distribution and the normal distribution etc.

We can check in accordance with the parameters received from the statistics. Generally we conduct the grid hierarchical query with the method from top to downside. The general steps of the algorithm are as follows:

1. Division level, select one level as the starting point of query, generally we select the level with the fewest units;
2. For the unit mesh of the selected level, we calculate and check for the related value of attributes;
3. From the attribute value of calculation and its restricted conditions, we mark each of the unit mesh as related or non-related;
4. If this level is the bottom level, then switch to step6, otherwise we will switch to step5;
5. We switch from the level structure to the unit mesh of lower level and conduct the calculation in accordance with step2;
6. The result of query is satisfied, switch to step8, otherwise switch to step7;
7. Recover the data to the related unit mesh for further process till the result is satisfied, switch to step8;
8. stop.

3.3 The thinking of R language programming for the algorithm of STINT

1. Read the sample data, process the data in need of research to the matrix format X,Y.
2. Set the mesh parameters as M, N

\[ A = \max(x) - \min(x) \]  \hspace{1cm} (3)
\[ B = \max(y) - \min(y) \]  \hspace{1cm} (4)
\[ r = \frac{A}{B} \]  \hspace{1cm} (5)
\[ M = \lceil \frac{(r + 1)^2 + 4 * (r * (n - 1)) - (r + 1)}{2} \rceil \]  \hspace{1cm} (6)
\[ N = \lceil \frac{M}{r} \rceil \]  \hspace{1cm} (7)

Store the sample data to the matrix of l mg

\[ \text{dex} = \frac{A}{4M} \hspace{1cm} \text{dey} = \frac{B}{4N} \]  \hspace{1cm} (8)
\[ xx = \text{seq}(\min(x) - \text{dex}, \max(x) + \text{dex}, \text{dex}) \]  \hspace{1cm} (9)
\[ yy = \text{seq}(\min(y) - \text{dey}, \max(y) + \text{dey}, \text{dey}) \]  \hspace{1cm} (10)

definition: \[ \text{Im} \text{g} = \text{matrix}(0, \text{nrow} = \text{length}(yy) - 1, \text{ncol} = \text{length}(xx) - 1) \]  \hspace{1cm} (11)
\[ m = \lceil \frac{x[i] - xx[1]}{\text{dex}} \rceil, (i = 1, 2, ..., n) \]  \hspace{1cm} (12)
\[ n = \lceil \frac{y[i] - yy[1]}{\text{dey}} \rceil, (i = 1, 2, ..., n) \]  \hspace{1cm} (13)
\[ \text{Im} \text{g}[n, m] = \text{Im} \text{g}[n, m] + 1 \]  \hspace{1cm} (14)

4. Transform the matrix unit of l mg as bw matrix

\[ \text{mn} = \text{dim}(	ext{Im} \text{g}) \]  \hspace{1cm} (15)
\[ \text{bw} = \text{matrix}(0, \text{mn}[1], \text{mn}[2]) \]  \hspace{1cm} (16)
\[ \text{Im} \text{g}[i, j] \geq 1, \text{bw}[i, j] = 1(i = 1, 2, ..., \text{mn}[1], j = 1, 2, ..., \text{mn}[2]) \]  \hspace{1cm} (17)

5. Conduct the mesh clustering

definition: \[ L = \text{BWLABEL}(\text{bw}, n) \]  \hspace{1cm} (18)
4. Empirical analysis

4.1 The analysis on the K-means clustering demonstration
With the traditional K-means clustering analysis, the initial classification number is set as 4. The clustering result is shown in the following Fig 2:

4.2 The process of data by DBSCAN clustering
Use the density-based algorithm based on DBSCAN that can eliminate the data of noise to classify all the observed data into 5 classes. The clustering result is indicated in the following Fig 3:

4.3 The process of data by STING clustering
According to the algorithm of STING clustering, classifying the observed data into 7 classes, because two of the classes have very few numbers, therefore, it shall be treated in accordance with the noise. The clustering result is shown in the following Fig 5:
4.4 Change of gas occupancy among categories
After clustering analysis, the sample data that does not meet the requirements of clustering will be ignored. The number of data of K-means clustering, DBSCAN clustering, STING clustering results are 639, 369 and 614 groups. Therefore, it is concluded that K-means clustering does not ignore any data, so there are too many data ignored by DBSCAN clustering, and STING clustering ignores very little data.

From Fig 3, we can see that the category change trend of STING algorithm is increasing. The shift free occupancy rate is increasing trend, with smaller offset and larger offset, occupancy rate is decreasing. In contrast (a) (b) diagram, STING grid clustering is more hierarchical than the category of DBSCAN density clustering.

4.5 The intensity of clustering
Given a group of data X, output the result $c_1, c_2, \ldots, c_c$ for the clustering.

The intensity of clustering is:

$$Cmp = \frac{1}{c} \sum_{c_i} \frac{Var(c_i)}{Var(X)}$$  \hspace{1cm} (19)

Of which c is the number of clustering, the class inside each clustering should be as close as possible, therefore, the density of the clustering is the smaller, the better.
Table 1. The denseness of three kinds of clustering

| type | K.means clustering | DBSCAN clustering | STING clustering |
|------|--------------------|-------------------|------------------|
| X1   | 0.1889403          | 0.0408644         | 0.07423269       |
| X2   | 0.2557622          | 0.05727167        | 0.1308406        |

The result shows that: the intensity of DBSCAN clustering is the best, the intensity of STING clustering is the second, the intensity of K-means clustering is the worst.

5. Conclusion

Through the comparison of three methods, we can know that the algorithm of STING integrates the advantages of K-means and dbscan, we need not determine the classifications in advance, we can find the clusters of any shape and can deal with any big data. Its intensity is good, the storage of information is comparatively independent, it is convenient for the query and the efficiency is high. Therefore, we can establish the highly effective clustering model for the data at the gas flow center of blast furnace, analyze the relationship between the distribution at the gas flow center and the gas shares and provide the reliable data analysis and guidance for the analysis on the furnace condition of blast furnace.

Foundation Item:
Item Sponsored by National Natural Science Foundation of Inner Mongolia China (2017MS(LH)0104), Innovation Fund Project of University of Science and Technology of Inner Mongolia(2015QDL17).

Biography
Ming-xin Li, Doctor, lecturer; E-mail: mingxin1581@163.com; 18504721581

References
[1] Han J, Kamber M. Data Mining: Concepts and Techniques. Morgan Kaufmnn Publishers,2001.
[2] Wang W,Yang J, Muntz R. STING: A Statistical Information Grid Approach to Spatial Data Mining. In: Proceedings of the 23rd VLDB Conference. Athens, Greece,1997.186-195.
[3] Hartigan J.A K-means clustering algorithm[J].Applied Statistics,1979,28:100-108.
[4] Ester M, Kriegel H-P,Sander J,et al. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In: Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining. 1996.266-231.
[5] Dong X F, Pinson D, Zhang S J, et al. Gas-powder flow in blast furnace with different shapes of cohesive zone [J]. Applied Mathematical Modeling, 2006, 11(30):1293-1309.
[6] Mats N, Henrik S. Prediction of a blast furnace burden distribution variable [J]. ISIJ International, 1996, 36(9):1142-1150.
[7] Timmer R, Droog J, Flierman G, et al. Radial gas distribution in the blast furnace top [J]. Steel’ Reasearch, 1997, 68(2):47-53.
[8] Jimenez J, Mochon J, De Ayala J S. Mathematical Model of Gas Flow Distribution in a Scale Model of a Blast Furnace Shaft [J]. ISIJ International, 2004, 44(3):518-526.
[9] PETTERSSON F, SAXEN H,DEB K. Genetic algorithm-based multicriteria optimization of I ronmaking in the blast furnace [J]. Materials and Manufacturing Process,2009,24,(3):343-349.
[10] Park, J., Bin, A. and Reis-Brod, F.P., Drying of pear d’Anjou with and without osmotic dehydration. Journal of Food Engineering, 56, pp. 97-103 2002.
[11] Marabi, A., Livings, S., Jacobsons, M. and Saguy, I.S., Normalized weibull distribution for modeling rehydration of food particulates. European Food Research Technology, 217, pp. 311-318, 2003. http://dx.doi.org/10.1007/s00217-003-0719-y.