REVIEW OF COMPUTER ENGINEERING STUDIES

ISSN: 2369-0755 (Print), 2369-0763 (Online)
Vol. 4, No. 1, March 2017, pp. 17-21
DOI: 10.18280/rces.040104

A publication of IIETA
International Information and Engineering Technology Association
Advancing the World of Information and Engineering
http://www.iieta.org/Journals/RCES

A power system active power network loss based calculation method on partial priority clustering algorithm

Bailin Liu[1], Xingwei Xu[2]

[1]School of Electrical & Electronic Engineering, North China Electric Power University, Beijing 102206, China
[2]State Grid Corporation of China, Northeast Division, Shenyang 110181, China

Email: liubailin0424@163.com

ABSTRACT

In this paper, we propose a grid active power loss calculation method based on partial priority clustering, which improves the calculation efficiency by utilizing the partial priority clustering algorithm and accurately divides the grid operation modes by fully utilizing the efficient clustering attribute of the correspondence between dataset of grid operation modes and grid loss values. The method proposed in this paper can analyze and process the large datasets accumulated during the long-term operation of the power grid and effectively perform evaluation on the grid active power loss. Results of grid simulation show that the calculation accuracy of this method is much higher than the traditional grid loss evaluation method.

Keywords: Grid Planning, Excitation System Adjustment Coefficient, Reactive Compensation.

1. INTRODUCTION

Grid active power losses refer to the power consumption and losses that occur during the energy transmission and sales process. Line loss rate, as an important technological and economic indicator of power enterprises, provides an important reference for the power losses of power suppliers at all levels in the grid. How to reduce power loss and utilize energy more effectively is an important issue that power enterprises are working on.

Grid active power loss can be obtained through the load flow calculation of the data of operation modes in different time periods. However, over time, the data of grid operation modes are greatly increased, leading to too much computation. In engineering practice, usually the typical operation mode is selected for load flow calculation, and then a coefficient is determined based on the operating experience to calculate the grid power loss [1-2], but in the operation process, the grid parameters, load and startup mode are always changing, so the results obtained by the above method are very different from the actual power losses.

References [3-4] calculated the grid loss probability distribution function by calculating the relationship between grid loss and the nodal injection. Reference [5] established a system state simulation model, which generates the data of multiple operation modes through sampling based on the Monte Carlo method and determines the probability distribution of grid losses through load flow calculation. This method has high calculation accuracy, but when the power grid scale is enlarged and huge amount of data are accumulated through long-term operation, the calculation efficiency will be greatly reduced, so this method is not suitable for long-term analysis and calculation.

As one of the main research methods for data mining, cluster analysis has been extensively applied in the new energy power generation, power consumption mode recognition and load forecasting [6-11]. In this paper, regarding the operation datasets of power grids, we propose a grid active power loss evaluation method based on partial priority clustering analysis. By searching for typical points, gradually finding each class and deleting the classified data in time, it simplifies the complexity of datasets, which increases the operation speed of the algorithm. We also apply the algorithm to the actual grid to test its effectiveness. From the calculation results, it can be seen that the algorithm can efficiently analyze operation datasets and take into account the main factors that affect the grid active power loss so that it can correctly obtain the typical operation mode for the power grid and accurately evaluate the grid active power loss.

2. PARTIAL PRIORITY CLUSTERING ALGORITHM

2.1 Basic concept

Definition 1. r-neighborhood definition: the *r*-neighborhood of a data object is an area with the object as the center and r as the radius.

Definition 2. Typical sample: if the r-neighborhood of an object in the sample contains at least *Minpts* objects, then this sample is called a typical sample.

Definition 3. Cluster center: the average value of all objects in a typical sample is called the cluster center. If Sample P is a typical sample, then,

$$C1 = \frac{\sum_{x_i \in P} x_i}{|P|} \quad (1)$$

(where i is the number of samples contained in Sample |P|) is the cluster center.

2.2 Partial priority clustering algorithm

Partial priority clustering algorithm consists of the following main steps:

(1) It first selects one sample *A* randomly from a dataset, chooses a point in A as the initial point and at the same time sets *Minpts* and *r*;

(2) If this point is not a typical sample, then it repeats Step (1) until finding the typical one;

(3) If the data amount in the r-neighborhood of this point is greater than *Minpts*, then A is a typical sample. In this case, it calculates the cluster center C_1 according to Formula (1);

(4) It starts clustering with C_1 as the center. If the distance between C_1 and an object xi is reduced to a certain value r, then it classifies this object into this type; otherwise, it makes judgment on the next data until it traverses all data in the dataset;

When the first type T_1 is created, it deletes all data falling within this type from the dataset to be clustered so that all these type T_1 data will not participate in the next classification to reduce the data complexity.

(5) It repeats Step (4) to classify the residual data.

The flow chart for this algorithm is shown in Figure 1.

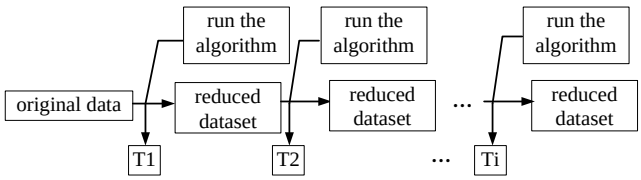

Figure 1. Partial priority clustering algorithm process

2.3 Advantages of the partial priority clustering algorithm

Compared with traditional clustering algorithms, the calculation based on partial priority clustering is more efficient and saves a lot of operation time:

(1) First, in order to improve the computational efficiency of typical samples and cluster centers, it uses random samples instead of the whole dataset, and only performs cluster center calculation for once, which greatly improves the computational efficiency;

(2) It chooses the average value of a typical sample as the cluster center, which is more representative and more suitable as the center of clustering. This can attract as much as possible data, so that the residual data set is reduced to the greatest extent;

(3) Each time a type is created, the data of this type will be deleted from the dataset to be clustered, reducing the complexity of the dataset.

As the partial priority clustering algorithm first finds the typical sample and then calculates the typical point, it is not sensitive to the input sequence of data. And as this algorithm looks for typical samples in the dense area, it is very sensitive to abnormal data and can determine whether a point is abnormal or not. Judging from the computation results, this algorithm can also deliver satisfactory result from multi-dimensional data.

3. CLUSTER FUSION OF THE CLUSTERING PARTIAL PRIORITY CLUSTERING ALGORITHM

Traditional clustering methods cannot deal with massive data. The clustering algorithms adopted often deliver blind and unstable classification results. To improve the typicality and stability of cluster centers, different attributes are given different weighs according to the effects of the attributes on the grid active power loss, and then use the cluster fusion algorithm to improve the algorithm accuracy so as to significantly improve the robustness and stability of the clustering algorithm.

3.1 Basic Concept

The cluster fusion technology [12-16] produces clustering members (sample data produced by different algorithms or different initialization and parameters in one algorithm) by integrating the existing clustering algorithms, and then "clusters" these members to maximize the information sharing of existing clustering results so as to be less sensitive to the original data distribution and deliver better results than a single algorithm. The cluster fusion technology can greatly improve the robustness and stability of the clustering algorithms and is able to perform parallel computing.

3.2 Steps of cluster fusion

For one dataset X= $\{x_1, x_2, \ldots, x_n\}$, where xi= $\{x_{i1}, x_{i2}, \ldots, x_{id}\}$, first we use the partial priority clustering algorithm to classify it. The feature of the partial priority algorithm is to first produce the first class, so when the first class is produced in each classification, a fusion method will be used to generate the final first class. Let's suppose the first class produced in the first classification is C_{11}, and the corresponding cluster center is C_{11}; the first class produced in the second classification is C_{12}, and the corresponding cluster center is C_{12}; and the first class produced in the *r*-th classification is C_{1r}. Among these $_r$ classifications, the data from the dataset are input in different sequences to make sure the first class will not be the same.

When the first class is produced in each classification, the first classes will be fused to generate the final first class. The fusion method used in this paper is to calculate the cluster center of the first class on a weighted basis with the ratio between the data amount of the class and total amount as the weight, as expressed in the following formula:

$$V_1 = \frac{|P_{11}|}{\sum_{i=1}^{r}|P_{1i}|}C_1^1 + \frac{|P_{12}|}{\sum_{i=1}^{r}|P_{1i}|}C_1^2 + \cdots \frac{|P_{1r}|}{\sum_{i=1}^{r}|P_{1i}|}C_1^r \quad (2)$$

where, V1 is the cluster center of the final first class, and |P|(i= 1, 2, …, k; j= 1, 2, …, r) stands for the amount of data

contained in Class Cij. After the cluster center is determined, establish an empty set T1, put V_1 into T_1, cluster the data with V1 as the center, and set the threshold value r. For data concentrated at any point x_i(i= 1, 2, …, n), if $|V_1-x_i|<r$, assign xi to the set T1. After all the points in the dataset are traversed, then T1 is the first class finally produced. After the first class is produced, delete the data contained in T_1 from each classification to reduce the complexity of data. Then produce the remaining k-1 classes by following the same steps.

4. GRID ACTIVE POWER LOSS EVALUATION METHOD BASED ON PARTIAL PRIORITY CLUSTERING

4.1 Obtaining the typical operation mode

The focus of this section is to study how to take the characteristics of grid operation data into full account to obtain the optimal typical grid operation mode cluster and at the same time ensure good clustering computation efficiency.

Factors affecting the grid active power loss include the generator startup mode, node load power and grid structural parameters. If the above data are used as the clustering attributes only, the clustering attributes corresponding to the output data of each node will increase with the enlarging scale of grid nodes. When this method is applied to a large power grid, problems like high-dimensional clustering will appear and the importance of total load and tie line power data in the grid loss evaluation will be drowned out, which will mislead the clustering. Therefore, it is necessary to consider the different effects of different attributes on the active power loss in the clustering process.

The grid active power loss evaluation method based on partial priority clustering takes the grid operation data and the running status parameters as the data samples, first adopts the partial priority clustering and cluster fusion methods to obtain the typical operation mode and the probability of occurrence and then evaluate the grid active power loss. The steps are as follows:

Step 1, take out the grid operation parameters at a certain moment from the grid operation data, use the generator startup mode, grid structural parameter and node load power, etc. that affect the grid active power loss greatly as the data samples at that moment, and then establish the operation dataset needed in the grid loss evaluation;

Step 2, use the partial priority clustering algorithm mentioned in Section 1.2 to extract the first class. Use the following formula to calculate the distance between data:

$$\begin{aligned}\left|X_i - X_j\right| &= K_m\left|x_{iGm} - x_{jGm}\right| + \\ &K_n\left|x_{iLn} - x_{jLn}\right| + K_w\left|x_{iNw} - x_{jNw}\right|\end{aligned} \tag{3}$$

where, $\left|X_i - X_j\right|$ is the distance between two data; x_{iGm} and x_{jGm} are the vectors formed by the generator startup mode and the generation power of the two data. $\left|x_{iGm} - x_{jGm}\right|$ is the distance between the two data; x_{iLm} and x_{jLm} are the vectors formed by the node load power of the two data. $\left|x_{iLn} - x_{jLn}\right|$ is the distance between the two; x_{iNw} and x_{jNw} are the vectors formed by the grid structural parameter of the two data and $\left|x_{iNw} - x_{jNw}\right|$ is the distance between the two; K_m, K_n and K_w are the weights of the terms.

The purpose of this paper is to evaluate the grid active power loss. The distance between the data should reflect the similarity of the grid operation modes at two different time, so data will be clustered according to the attributes that have large effects on the active power loss of the grid. In addition, according to the theory of power flow calculation, different data attributes affect the grid loss differently, so in the distance calculation, different weights are assigned to the various attributes to ensure that the calculated results of distance can be used to effectively distinguish between different modes of operation.

Step 3, according to the cluster fusion method in Section 2.2, adjust the input sequence of the power grid operation data, perform clustering for multiple times, and fuse the first class in each classification to get the final first class and calculate the cluster center; classify all points from the power operation dataset whose distance to the center of the first class is less than a fixed value as the first class. After all data are traversed, we get the first complete cluster T1, and delete all data in T1 from the grid operation dataset;

Step 4, repeat Step 2 and 3 to produce the remaining k-1 clusters by the same method.

4.2 Evaluation on the grid active power loss

The k clusters obtained by the method described in the above section are the typical operation modes. The proportion of the data amount contained in each cluster in the total data amount is the frequency of occurrence. The attributes of the cluster center data are the grid operation parameters of typical operation modes.

Through the mode flow calculation of all typical operation modes, we can obtain the evaluation results of the grid active power loss from the following formula.

$$P_{Loss} = \sum_{i=1}^{k} \lambda_i P_{lossi} \tag{4}$$

where, P_{Loss} is the total grid power loss, λ_i is the frequency of occurrence of the i-th operation mode, and P_{lossi} is the active power loss in the i-th operation mode.

5. EXAMPLE ANALYSIS

In order to illustrate the effects of the proposed method, we apply two grid active power loss calculation methods to an actual grid in some area. The power grid operation dataset consists of the SCADA data of the grid from 2012~2013. The data are acquired at an interval of 1 hour. The calculation results of different schemes are compared as follows.

Scheme 1 takes the one-dimensional data formed by generator startup mode, generation power, node load power and the on-off state of the power transmission lines from the grid operation data as the attribute data of the operation mode at the moment, and uses traditional clustering methods to perform clustering to obtain the typical operation modes and

the frequency of occurrence, and then calculates the total grid power loss using Formula (4).

In Scheme 1, the heterogeneous attribute data clustering algorithm [] is adopted, which does not consider the different effects of different data attributes on the power loss in the calculation process. Table 1 shows the typical operation modes and the frequency of occurrence.

From Table 1 and 2, it can be seen that Scheme 1, which uses the traditional heterogeneous attribute data clustering algorithm to determine the typical operation modes, cannot effectively extracts key attribute data from multidimensional arrays; under the method proposed in this paper, there are more clusters of typical operation modes, indicating that this method can fully utilize the grid operation dataset and explore the differences between different grid operation statuses and the typical operation modes obtained are more representative.

Calculation results of the total grid power loss are listed in Table 3.

From Table 1 and 2, it can be seen that Scheme 1, which uses the traditional heterogeneous attribute data clustering algorithm to determine the typical operation modes, cannot effectively extracts key attribute data from multidimensional arrays; under the method proposed in this paper, there are more clusters of typical operation modes, indicating that this method can fully utilize the grid operation dataset and explore the differences between different grid operation statuses and the typical operation modes obtained are more representative.

Table 2. The typical operation modes obtained and the frequency of occurrence

Cluster center	Clustered data volume	Frequency of occurrence
9	432	2.47%
25	264	1.51%
39	408	2.33%
44	648	3.70%
89	432	2.47%
135	480	2.74%
144	432	2.47%
181	792	4.52%
199	336	1.92%
234	552	3.15%
278	456	2.60%
290	1272	7.26%
303	504	2.88%
322	552	3.15%
323	408	2.33%
359	312	1.78%
375	960	5.48%
434	432	2.47%
452	696	3.97%
464	168	0.96%
492	408	2.33%
511	1464	8.36%
518	600	3.42%
565	552	3.15%
575	240	1.37%
595	480	2.74%
621	552	3.15%
664	432	2.47%
676	984	5.62%
692	432	2.47%
697	312	1.78%
717	528	3.01%

Calculation results of the total grid power loss are listed in Table 3.

Table 3. Comparison of grid power loss evaluation results

	Grid power loss (10,000kw)	Error
Actual grid power loss	17644.85	
Calculated result in Scheme 1	19952.80	13.08%
Calculated result in the proposed method	18269.48	3.54%

From Table 1 and 2, it can be seen that Scheme 1, which uses the traditional heterogeneous attribute data clustering algorithm to determine the typical operation modes, cannot effectively extracts key attribute data from multidimensional arrays; under the method proposed in this paper, there are more clusters of typical operation modes, indicating that this method can fully utilize the grid operation dataset and explore the differences between different grid operation statuses and the typical operation modes obtained are more representative.

Calculation results of the total grid power loss are listed in Table 3.

6. CONCLUSIONS

In this paper, we propose a grid active power loss evaluation method based on partial priority clustering. Under this method, each time a cluster is created, the data of this cluster will be deleted from the dataset to reduce the complexity of the dataset and improve computational efficiency; during the clustering, different weights will be assigned to different attribute data according to their effects on the active power loss, so that the clustering result of the grid operation modes will better reflect the different grid operation statuses. Results of grid simulation show that when the method proposed in this paper is applied to evaluate the grid active power loss, the calculation accuracy is improved and the error is significantly reduced.

ACKNOWLEDGMENTS

Sponsored by the key project of national natural science foundation of China (51437003).

REFERENCES

[1] Li L.C., Wang J.Y., Chen L.Y. (1999). Optimal reactive power planning of electrical power system, *Proceedings of the CSEE*, Vol. 19, pp. 66-69.

[2] Rao J., Zhang Y.J., Ren Z. (2004). Study of optimal reactive power planning for urban network, *East China Electric Power*, Vol. 32, No. 11, pp. 734-737.

[3] Nadira R., Wu F.F., Maratukulam D.J. (1993). Bulk transmission system loss analysis, *IEEE Trans on Power System*, Vol. 8, No. 2, pp. 405-416. DOI: 10.1109/59.260846

[4] Mu G., Zhou Y. (1998). Computation and analysis of probabilistic distribution function of power loss in a

bulk transmission system, *Proceedings of the CSEE*, Vol. 18, No. 6, pp. 426-428.

[5] Yan W., Lv Z.S., Li Z.J., Long X.P., Yang X.M. (2007). Monte Carlo simulation and transmission loss evaluation with probabilistic method, *Proceedings of the CSEE*, Vol. 27, No. 34, pp. 39-45.

[6] Zhao Y.N., Ye L., Zhu Q.W. (2014). Characteristics and processing method of abnormal data clusters caused by wind curtailment in wind farms, *Automation of Electric Power System*, Vol. 38, No. 21, pp. 39-46. DOI: 10.7500/AEPS20131213010

[7] Lin J.K., Liu L., Zhang W.B., Liu J., Wang G. (2013). Load modeling and parameter identification based on random fuzziness clustering, *Automation of Electric Power System*, Vol. 37, No. 14, pp. 50-58.

[8] Zhang L.Z., Wang Q., Shu J. (2011). A composite generation-transmission system reliability assessment method based on clustering of optimal multiplier vector, *Automation of Electric Power System*, Vol. 35, No. 6, pp. 14-19.

[9] Peng X.G., Lai J.W., Chen Y. (2014). Application of clustering analysis in typical power consumption profile analysis, *Power System Protection and Control*, Vol. 42, No. 19, pp. 68-73.

[10] Wang D.W., Sun Z.W. (2015). Big data analysis and parallel load forecasting of electric power user side, *Proceedings of the CSEE*, Vol. 35, No. 3, pp. 527-537.

[11] Zhang S.X., Zhao B.Z., Wang F.Y. (2015). Short-term power load forecasting based on big data, *Proceedings of the CSEE*, Vol. 35, No. 1, pp. 37-42.

[12] Zhang D., Yang L.Y., Wang W.Y. (2005). Clustering ensemble approaches: overview, *Computer Application Research*, No. 12, pp. 8-10.

[13] Wu X.X., Ni Z.W., Ni L.P. (2012). Clustering ensemble algorithm based on fractal, *Journal of Jilin University (Engineering and Technology Edition)*, Vol, 9, No. 42, pp. 364-367. DOI: 10.13229/j.cnki.jdxbgxb2012.s1.060

[14] Zhao Y., Li B., Li X., Liu W.H., Ren S.J. (2006). Cluster ensemble method for databases with mixed numeric and categorical values, *Journal of Tsinghua Uni (Sci & tech)*, Vol. 46, No. 10, pp. 1673-1676.

[15] Yang C.Y., Zhou J. (2007). A heterogeneous data stream clustering algorithm, *Chinese Journal of Computer*, Vol. 30, No. 8, pp. 1365-1371.

[16] Wu W.L., Liu Y.S., Zhao J.H. (2009). New mixed clustering algorithm, *Journal of System Simulation*, Vol. 19, No. 1, pp. 16-18.