

# Fault Detection Based on Hierarchical Cluster Analysis in Wide Area Backup Protection System

Yagang ZHANG, Jinfang ZHANG, Jing MA, Zengping WANG

*Key Laboratory of Power System Protection and Dynamic Security Monitoring and Control under Ministry of Education,*

*North China Electric Power University, Baoding, China*

*Email: yagangzhang@gmail.com*

**Abstract:** In wide area backup protection of electric power systems, the prerequisite of protection device's accurate, fast and reliable performance is its corresponding fault type and fault location can be discriminated quickly and defined exactly. In our study, global information will be introduced into the backup protection system. By analyzing and computing real-time PMU measurements, basing on cluster analysis theory, we are using mainly hierarchical cluster analysis to search after the statistical laws of electrical quantities' marked changes. Then we carry out fast and exact detection of fault components and fault sections, and finally accomplish fault isolation. The facts show that the fault detection of fault component (fault section) can be performed successfully by hierarchical cluster analysis and calculation. The results of hierarchical cluster analysis are accurate and reliable, and the dendrograms of hierarchical cluster analysis are in intuition.

**Keywords:** wide area backup protection, phasor measurement unit, PMU, wide area measurement system, WAMS, fault detection, cluster analysis

## 1 Introduction

Electric power system is one of the most complex artificial systems in this world, which safe, steady, economical and reliable operation plays a very important part in guaranteeing socioeconomic development, even in safeguarding social stability. In early 2008, the infrequent disaster of snow and ice that occurred in the south of China had confirmed it again. The complexity of electric power system is determined by its characteristics about constitution, configuration, operation, organization, etc., which has caused many disastrous accidents, such as the large-scale blackout of America-Canada electric power system on August 14, 2003, the large-scale blackout of Chinese Hainan electricity grid on September 26, 2005. In order to resolve this difficult problem, some methods and technologies that can reflect modern science and technology level have been introduced into this domain, such as computer and communication technology, control technology, superconduct and new materials technology and so on. Obviously, no matter what we adopt

new analytical method or technical means, we must have a distinct recognition of electric power system itself and its complexity, and increase continuously analysis, operation and control level [1-3].

Relay protection is the first line of guaranteeing large-scale electricity grid's safety. The faults in electric power system are inevitable. If protection devices can operate rightly, quickly and reliably, the deterioration of system status will be checked effectively, then it will play a decisive role to protect electricity grid's safe operation. Otherwise, it will accelerate system crashes, as a result, large-scale and long-time power blackout will continue. After counting seventeen years accident data in electric power system, North American Electric Reliability Council (NERC) has found: 63% accidents in electric power system are concerned with the incorrect operation of relay protection. The large-scale power blackouts occurred in China and other countries of the last thirty years have also indicated: the large-scale power blackout accidents are often raised from the improper cooperation

or chain reaction of protection devices. The large-scale blackout of America-Canada electric power system was just because the removal of four connection lines between Akron and Cleveland in northern Ohio by backup protection for overload, and the accident spread rapidly. The backup protection in current electricity grid is only reflecting the information of protection installation position, which will be affected by topological connecting relations and operation modes. In order to guarantee its reliability, we can only carry through configuration and setting according to the most rigorous condition. In order to guarantee its selectivity, we have to sacrifice the rapidity and sensitivity of backup protection [4][5]. In recent years, the appearance of wide area measurement system (WAMS) affords the possibility for introducing system information into backup protection system. WAMS can obtain synchronously electrical measurements in the whole power system, and realize power system dynamic process monitoring and control. It can also decrease the update speed of measurements from seconds to tens of millisecond, and create condition to realize power system dynamic process control, which will help us carry through backup protection design based on global optimal angles of electricity grid, and afford the possibility for resolving dynamic security monitoring, control and protection of complex large-scale electricity grid.

When electric power system operates from normal state to failure or abnormal operates, its electric quantities (current magnitude, voltage magnitude and their angles, etc.) may change significantly. In our researches, global information will be introduced into the backup protection system. After some accidents, utilizing real-time measurements of phasor measurement unit (PMU) [6-10], basing on multivariate statistical analysis theory [11-13], we are using mainly cluster analysis technology [14-19], and seeking after for statistical laws of electrical quantities' marked changes. Then we can carry out fast and exact detection of fault components and fault sections, and hereby ascertain protection com-

ponents associated with them. Finally we can accomplish fast and exact fault isolation.

The cluster analysis theory is one of multivariate statistical analysis theory, which is a synthetical analysis theory. In recent years, as the development of computer application technology and the demand of scientific research and production, multivariate statistical analysis theory has been applied successfully to many researches of various fields, such as geology, weather, hydrology, iatrolgy, industry, agriculture, and economy, etc. It has been an efficient theory that can resolve different kinds of complex problems. Basing on statistical theory, we have carried out large numbers of basic researches in nonlinear dynamical systems [20-22]. In this paper, we are using mainly cluster analysis of multivariate statistical analysis theory to resolve fault detection problem in wide area backup protection of electric power systems.

## 2 Cluster Analysis Theory

Theories of classification come from philosophy, mathematics, statistics, psychology, computer science, linguistics, biology, medicine, and other areas. Cluster analysis can also be named classification, which is concerned with researching the relationships within a group of objects in order to establish whether or not the data can be summarized validly by a small number of clusters of similar objects. That is, cluster analysis encompasses the methods used to:

- Identify the clusters in the original data;
- Determine the number of clusters in the original data;
- Validate the clusters found in the original data.

Cluster analysis has great strength in data analysis and has been applied successfully to the researches of various fields.

Suppose there are  $n$  samples, each sample has  $m$  indexes (variables), the observation data can be expressed as,

$$x_{ij} \quad (i = 1, \dots, n, \quad j = 1, \dots, m)$$

In these data, the definition of mean is:

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad (j = 1, 2, \dots, m)$$

the definition of standard deviation is:

$$S_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2} \quad (j = 1, 2, \dots, m).$$

### 2.1 The Distance and Similar Coefficient Between Samples

The most commonly used measurement that describes the degree of relationship is distance,  $d_{ij}$  is usually denoted the distance between samples  $X_{(i)}$  and  $X_{(j)}$ , the general demands are:

- (1).  $d_{ij} \geq 0$ , for arbitrary  $i, j$ , and  $d_{ij} = 0 \Leftrightarrow X_{(i)} = X_{(j)}$ ;
- (2).  $d_{ij} = d_{ji}$ , for arbitrary  $i, j$ ;
- (3).  $d_{ij} \leq d_{ik} + d_{kj}$ , for arbitrary  $i, j, k$  (Triangle inequality).

The distance definitions in common use include:

1) Minkovski distance

$$d_{ij}(q) = \left[ \sum_{t=1}^m |x_{it} - x_{jt}|^q \right]^{\frac{1}{q}} \quad (i, j = 1, 2, \dots, n)$$

2) Lance distance ( $x_{ij} > 0$ )

$$d_{ij}(L) = \frac{1}{m} \sum_{t=1}^m \frac{|x_{it} - x_{jt}|}{(x_{it} + x_{jt})} \quad (i, j = 1, 2, \dots, n)$$

This is a measure without dimension, and it is insensitive to big singular values.

3) Mahalanobis distance

$$d_{ij}(M) = (X_{(i)} - X_{(j)})' S^{-1} (X_{(i)} - X_{(j)}) \quad (i, j = 1, 2, \dots, n)$$

Hereinto,  $S^{-1}$  is an inverses matrix of samples' covariance matrix.

4) Oblique space distance

In order to overcome the influence of relativity be-

tween variables, one can define the distance of oblique space:

$$d_{ij} = \left[ \frac{1}{m^2} \sum_{k=1}^m \sum_{l=1}^m (x_{ik} - x_{jk})(x_{il} - x_{jl}) r_{kl} \right]^{\frac{1}{2}} \quad (i, j = 1, 2, \dots, n)$$

Hereinto,  $r_{kl}$  is the correlation coefficient between  $X_k$  and  $X_l$ .

### 2.2 The Similar Coefficient and Distance Between Variables

Suppose  $C_{ij}$  can be expressed as the similar coefficient between  $X_i$  and  $X_j$ , the general demands are:

- (1).  $C_{ij} = \pm 1 \Leftrightarrow X_i = aX_j$  ( $a \neq 0$ , constant);
- (2).  $|C_{ij}| \leq 1$ , for arbitrary  $i, j$ ;
- (3).  $C_{ij} = C_{ji}$ , for arbitrary  $i, j$ .

$|C_{ij}|$  close to one means that  $X_i$  and  $X_j$  have near relationship, otherwise,  $C_{ij}$  close to zero means that they have distant relationship. The similar coefficients in common use are included angle Cosine and correlation coefficient.

1) Included angle Cosine

These  $n$  observed values  $(x_{1i}, x_{2i}, \dots, x_{ni})$  of  $X_i$  can be regarded as vectors in  $n$ -dimensional space, and the angle  $\alpha_{ij}$ 's Cosine of  $X_i$  and  $X_j$  is called similar coefficient of these two variables, namely

$$C_{ij}(1) = \text{Cos}[\alpha_{ij}] = \frac{\sum_{t=1}^n x_{it} x_{jt}}{\sqrt{\sum_{t=1}^n x_{it}^2} \sqrt{\sum_{t=1}^n x_{jt}^2}} \quad (i, j = 1, 2, \dots, m)$$

2) Correlation coefficient

The correlation coefficient is just the included angle Cosine after the data have been standardized.  $r_{ij}$  is expressed in common use the correlation coefficient of  $X_i$  and  $X_j$ , here we define it as  $C_{ij}(2)$ ,

$$C_{ij}(2) = \frac{\sum_{t=1}^n (x_{it} - \bar{x}_i)(x_{jt} - \bar{x}_j)}{\sqrt{\sum_{t=1}^n (x_{it} - \bar{x}_i)^2} \sqrt{\sum_{t=1}^n (x_{jt} - \bar{x}_j)^2}} \quad (i, j = 1, 2, \dots, m)$$







Figure 6. Branch set around BUS-18 fault node

cluster analysis of fault component and non-fault component (fault section and non-fault section).

### 3.3 Fault Detection of IEEE39-Bus System Based on Node Positive Sequence Voltage

Likewise, we calculate the node positive sequence voltage at  $T_{-1}$ ,  $T_0$  (Fault) and  $T_1$  three times. Figure 5 is the dendrogram of hierarchical cluster analysis based on node positive sequence voltage.

In the hierarchical cluster analysis based on node positive sequence voltage, the fault characteristic of Bus-18 is very obvious. Bus-18, Bus-3 and Bus-17 can be regarded as a cluster. For Bus-3 and Bus-17 are directly connected with Bus-18, the fault of Bus-18 will undoubtedly affect its adjacent nodes, as the case stands, Bus-18, Bus-3 and Bus-17 have also constituted accurately the fault section. Figure 6 is the branch set around Bus-18 fault node. So, in accordance with three-phase short-circuit to ground fault, based on node positive sequence voltage, the fault location can be detected exactly by the hierarchical cluster analysis.

These instances have fully proven that fault detection of fault component (fault section) can be performed by hierarchical cluster analysis and calculation. The results of hierarchical cluster analysis are accurate and reliable, and the dendrograms of hierarchical cluster analysis are in intuition.

## 4 Conclusions and Discussion

In wide area backup protection of electric power systems, the prerequisite of protection device's accurate, fast and reliable performance is its corresponding fault type and fault location can be discriminated quickly and defined exactly. In our researches, global information has been introduced into the backup protection system, basing on cluster analysis theory, we are using mainly hierarchical

cluster analysis technology, and seeking after for statistical laws of electrical quantities' marked changes by analyzing and computing real-time PMU measurements, thereby we carry out fast and exact detection of fault components and fault sections, and finally accomplish fault isolation.

Multivariate statistical analysis theory is an efficient theory that can resolve different kinds of complex problems. It has been applied successfully to many researches of various fields, and can analyze statistical law contained within subject, even multi-object and multi-index are associated together. In this paper, we are using mainly hierarchical cluster analysis of multivariate statistical analysis theory to resolve fault detection problem in wide area backup protection of electric power systems, and have got some ideal results. In the study of electric power systems, multivariate statistical analysis theory must also have a good prospect of application.

## Acknowledgements

This research was supported partly by Key Program of National Natural Science Foundation of China (50837002) and the Science Foundation for the Doctors of NCEPU.

## REFERENCES

- [1] J. X. Yuan, "Wide area protection and emergency control to prevent large scale blackout," China Electric Power Press, Beijing, 2007.
- [2] L. Ye, "Study on sustainable development strategy of electric power in China in 2020," *Electric Power*, Vol. 36, No. 10, 1-72003.
- [3] Y. S. Xue, "Interactions between power market stability and power system stability," *Automation of Electric Power Systems*, Vol. 26, No. 21-22, pp. 1-6, 1-4, 2002.
- [4] Q. X. Yang, "A review of the application of WAMS information in electric power system protective relaying," *Modern Electric Power*, Vol. 23, No. 3, pp. 1, 2006.
- [5] J. Yi and X. X. Zhou, "A survey on power system wide-area protection and control," *Power System Technology*, Vol. 30, No. 8, pp. 7-13, 2006.
- [6] A. G. Phadke and J. S. Thorp, "Synchronized phasor measurements and their applications," Springer-Verlag, New York, 2008.

- [7] T. S. Bi, X. H. Qin, and Q. X. Yang, "A novel hybrid state estimator for including synchronized phasor measurements," *Electric Power Systems Research*, Vol. 78, No. 8, pp. 1343-1352, 2008.
- [8] C. Wang, C. X. Dou, X. B. Li, and Q. Q. Jia, "A WAMS/PMU-based fault location technique," *Electric Power Systems Research*, Vol. 77, No. 8, pp. 936-945, 2007.
- [9] C. Rakpenthai, S. Premrudeepreechacharn, S. Uatrongjit, and N. R. Watson, "Measurement placement for power system state estimation using decomposition technique," *Electric Power Systems Research*, Vol. 75, No. 1, pp. 41-49, 2005.
- [10] J. N. Peng, Y. Z. Sun, and H. F. Wang, "Optimal PMU placement for full network observability using Tabu search algorithm," *International Journal of Electrical Power & Energy Systems*, Vol. 28, No. 4, pp. 223-231, 2006.
- [11] X. Q. He, "Modern statistical analysis methods and applications," China Renmin University Press, Beijing, 2007.
- [12] X. L. Yu and X. S. Ren, "Multivariate statistical analysis," China Statistic Press, Beijing, 1998.
- [13] Y. T. Zhang and K. T. Fang, "Introduction to multivariate statistical analysis," Science Press, Beijing, 1982.
- [14] A. Z. Arifin and A. Asano, "Image segmentation by histogram thresholding using hierarchical cluster analysis," *Pattern Recognition Letters*, Vol. 27, No. 13, pp. 1515-1521, 2006.
- [15] X. Otazu and O. Pujol, "Wavelet based approach to cluster analysis: Application on low dimensional data sets," *Pattern Recognition Letters*, Vol. 27, NO. 14, pp.1590-1605, 2006.
- [16] H. S. Park and D. K. Baik, "A study for control of client value using cluster analysis," *Journal of Network and Computer Applications*, Vol. 29, No. 4, pp. 262-276, 2006.
- [17] V. Tola, F. Lillo, M. Gallegati, and R. N. Mantegna, "Cluster analysis for portfolio optimization," *Journal of Economic Dynamics and Control*, Vol. 32, No. 1, pp. 235-258, 2008.
- [18] W. X. Zhao, P. K. Hopke, and K. A. Prather, "Comparison of two cluster analysis methods using single particle mass spectra," *Atmospheric Environment*, Vol. 42, No. 5, pp. 881-892, 2008.
- [19] M. Templ, P. Filzmoser, and C. Reimann, "Cluster analysis applied to regional geochemical data: Problems and possibilities," *Applied Geochemistry*, Vol. 23, No. 8, pp. 2198-2213, 2008.
- [20] Y. G. Zhang, P. Zhang, and H. F. Shi, "Statistic character in nonlinear systems," *Proceedings of the Sixth International Conference on Machine Learning and Cybernetics*, Hong Kong, Vol. 5, pp. 2598-2602, 2007.
- [21] Y. G. Zhang, C. J. Wang, and Z. Zhou, "Inherent randomness in 4-symbolic dynamics," *Chaos, Solitons and Fractals*, Vol. 28, No. 1, pp. 236-243, 2006.
- [22] Y. G. Zhang and C. J. Wang, "Multiformity of inherent randomness and visitation density in  $n$ -symbolic dynamics," *Chaos, Solitons and Fractals*, Vol. 33, No. 2, pp. 685-694, 2007.