Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: http://www.science-gate.com/IJAAS.html

Cosine similarity cluster analysis model based effective power systems fault identification

Tan Yong Sing [1], Syahrel Emran Bin. Siraj [1], Raman Raguraman [1], Pratap Nair Marimuthu [1], K. Nithiyananthan [2,*]

[1]*Faculty of Engineering and Computer Technology, AIMST University, Bedong, Kedah, Malaysia*
[2]*Department of Electrical and Electronics Engineering, Karpagam College of Engineering, Coimbatore, India*

ARTICLE INFO

Article history:
Received 20 October 2016
Received in revised form
2 December 2016
Accepted 5 December 2016

Keywords:
Cluster analysis
Cosine similarity models
Power system transmission line faults
Data mining

ABSTRACT

The main objective of this paper is to develop a novel technique using Cluster Analysis with Cosine Similarity model to detect power system transmission lines fault and the types of fault that had occurred in power system. A test case of IEEE30 bus power system and different types of fault are simulated using PowerWorld v.18 software. Statistical Package for the Social Science (SPSS) software was used to implement Cluster Analysis with Cosine Similarity models towards the data simulated by PowerWorld software. The proposed model has two processes the first process will determine 3 phase fault, single line-to-ground fault and double line-to-ground fault. A Second process will determine line-to-line fault and double line-to-ground fault too. In some cases double line-to-ground fault can be determined in first process, but in this paper the double line-to-ground fault was determined by a second process. In the proposed model each phase of the nominal per unit bus voltages will be clustered and the output will be evaluated together with uninterrupted phase voltage data in order to determine the bus at fault and the types of fault. The innovative proposed model had successfully determined the bus at fault and the types of fault in 30 bus Power System.

1. Introduction

As the result of the growth of power system structure, each year, detection of power system transmission lines fault becomes a difficult task as the system becomes more complex and overflow with data in the data acquisition system database. By making use of the data acquired for detecting power system transmission lines fault will become a much easier task. Recently, Knowledge Discovery in Database (KDD) had gained popularity and advanced aggressively. Data mining is one of the processes in KDD where useful information and hidden knowledge are discovered from the raw data, from these knowledge discovery will help in doing making decision which power system bus is at fault and what type of fault that had occurred. Transmission lines play major roles in power system which enabled electrical power to transfer from the generating stations to the end consumer. Failure of transmission lines will cause power failure, which might cause financial damages, the safety of workers in industries and safety of normal consumer at risk. Hence, restoration of transmission, lines failure must be done expeditiously. Overhead transmission lines fault can be categorized into two categories Symmetrical faults and Asymmetrical faults are shown in Fig. 1. Three phase fault is a symmetrical fault where all three phases are affected equally sometimes it is also called as balanced fault. Asymmetrical unlike Symmetrical fault the faults are not affecting all the three phases equally. There are three types of Asymmetrical faults such as Single Line-to-Ground fault, Line-to-Line fault and Double Line-to-Ground fault.

Over the decades, electricity became so important in our modern world power system getting more advanced and intelligent in order to secure continuous electrical supply to consumer. Many researchers and experts had built a concrete foundation in transmission lines fault detection. Early models such as impedance-based which utilizing pure fault and pre-fault data to detect the transmission lines fault location (Takagi et al., 1982), travelling waves which utilizing high frequency electromagnetic waves (Rohrig, 1931). As a travelling waves model had a lot of noise which will affect the fault detection, Discrete Wavelet

* Corresponding Author.
Email Address: nithiieee@yahoo.co.in (K. Nithiyananthan)
https://doi.org/10.21833/ijaas.2017.01.018
2313-626X/© 2017 The Authors. Published by IASE.

Transform (DWT) was employed in travelling wave models to eliminate the distorted signals.

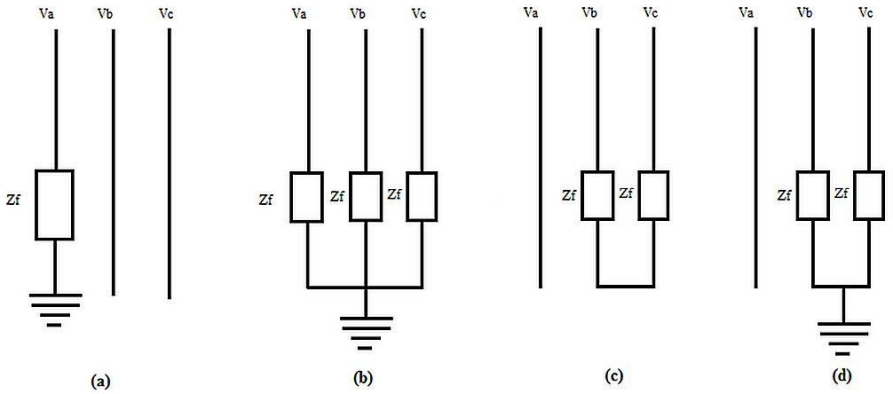

Fig. 1: (a) Single line-to-ground fault (b) three phase fault (c) line-to-line fault (d) double line-to-ground fault

Tabatabaei et al. (2012) included actual data acquisition features into DWT by using Global Positioning System (GPS). However, DWT models unable to detect line-to-line fault and single line-to-ground fault. The mother wavelet chosen in DWT must be chosen properly. More non-conventional models applied to power system fault detection such as Artificial Intelligent techniques. Dalstein and Kulicke (1995) had employed Artificial Neural Network (ANN) on power system transmission lines fault and classified the fault type, Surya et al. (2014) proposed to detect single line-to-ground fault using ANN. Ferrero et al. (1995) used Fuzzy Logic model to detect single line-to-ground fault and double line-to-ground fault. It further improves to include detection of line-to-line fault by Wang and Keerthipala (1998) using Fuzzy-Neuro approach. Increase in transmission lines fault types identification had proposed by Das and Reddy (2005) based on Fuzzy logic model. However, it is time consuming to detect transmission lines fault due to the large network of power system. A better solution or enhancement of non-conventional models is KDD techniques. Many researchers and experts very much interested in applying KDD models such as data mining techniques into field of power systems. Zhang et al. (2009) had proposed IEEE 9 bus transmission lines fault detection using Cluster Analysis based on Data Mining techniques. However, it was only tested for single line-to-ground fault and unable to detect various types of transmission lines faults. The objective of this paper is to propose a novel model using Hierarchical Clustering with Cosine Similarity model in determining symmetrical and asymmetrical transmission lines fault for IEEE30 bus power system. This paper is organized as follows, Section 2 briefly discussed on Cluster Analysis, Hierarchical Clustering techniques and Cosine Similarity models, Section 3 will discuss the proposed model using Hierarchical clustering techniques to identify the bus at fault and the types of fault that occur, Section 4 presents the results of simulations. Finally, Section 5 will be the concluding remarks. Section 6 is the acknowledgment for this research work. Now –a – days due to continuous expansion of Power System Network, controlling and monitoring of Power systems is unavoidable. Solutions through advanced data communications model are in evident (Mani et al., 2015; Nithiyananthan and Ramachandran, 2002).

2. Cosine similarity hierarchical cluster analysis

Hierarchical clustering (Jain et al., 1999) is one of the Cluster Analysis algorithms which represent the final result in a binary tree structure which is called dendrogram. There are few advantages against Partitional clustering such as by employing Hierarchical clustering it will provide the clustering information at each level of cluster and it does not need the user to randomly provide the number of clusters. Basically, there are two different approaches in Hierarchical clustering, which are Agglomerative (bottom-up) and Divisive (top-down). Agglomerative approach will start with one data in each cluster, hence if there are four data, then it start with four clusters then it will continue merging two clusters each level based on the similarity between clusters until single cluster was formed. Whereas, Divisive approach is the reverse of Agglomerative approach where it starts with the single cluster containing all data points and recursively splits to form the dendrogram (Fig. 2).

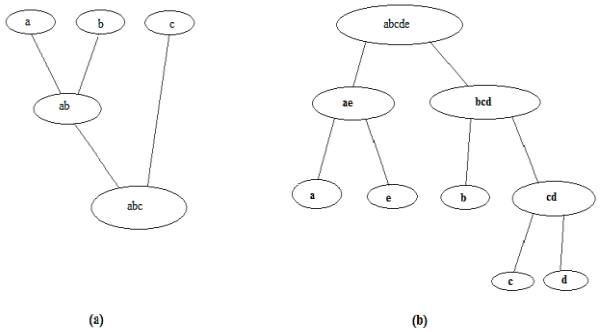

Fig. 2: (a) Agglomerative clustering (b) Divisive clustering

The crucial step in the Agglomerative algorithm is the right selections of models to compute the similarity or dissimilarity values based on the data attributes by treating the data as vectors. The transmission line fault data after subjected to data transformation perform very well with Cosine

Similarity model. Cosine Similarity model focused on the minimal angle between two vectors. Generally, it can be mathematical represented as (Eq. 1):

$$cos\theta = \frac{a^T b}{\|a\| \|b\|} \quad (1)$$

where, a and b are vectors.

Based on Cauchy-Schwarz inequality (Aldaz et al., 2015) stated that for all vectors x and y of an inner product space it is true that (Eqs. 2 and 3):

$$|a^T b| \leq \|a\| \cdot \|b\| \quad (2)$$

$$-1 \leq cos\theta \leq 1 \quad (3)$$

Therefore, from Eq. 3 the Cosine Similarity values are within the 1 and -1 range where 1 is 0 and -1 is 180. Cosine Similarity model and Agglomerative clustering technique were very suitable for our proposed model which improve the accuracy of fault detection towards the transformed raw data.

3. Proposed cosine similarity cluster analysis model

The detection of transmission lines fault become hard, especially todays power system networks are large and distributed. Precise and fast transmission lines fault identification will prevent huge losses in term of economy or society. Many of the conventional models discussed in Section 1 unable to perform very well due to the complexity of today power system networks and many noise data available. The proposed model utilized the three phase bus voltages (Va, Vb and Vc) at fault from Power World simulation of IEEE30 bus power system. The data collected was transformed into values 0 and about 1 at first and subjected to Agglomerative clustering with Cosine Similarity models for each phase voltages individually. If there are non-deterministic outputs from the dendrogram after evaluation, the raw data undergone different transformation equation at data transformation stage followed by clustering of each phase voltages. Hence, there are two main processes executed in sequence in order to determine all types of transmission lines faults. The first process identify three phase fault, single line-to-ground fault and double line-to-ground fault (which was not simulated in this paper), whereas, the second process determines line-to-line fault and double line-to-ground fault. All the identification of bus at fault and the type of fault was compared with the simulated data in order to make sure the result obtain was consistent (Fig. 3).

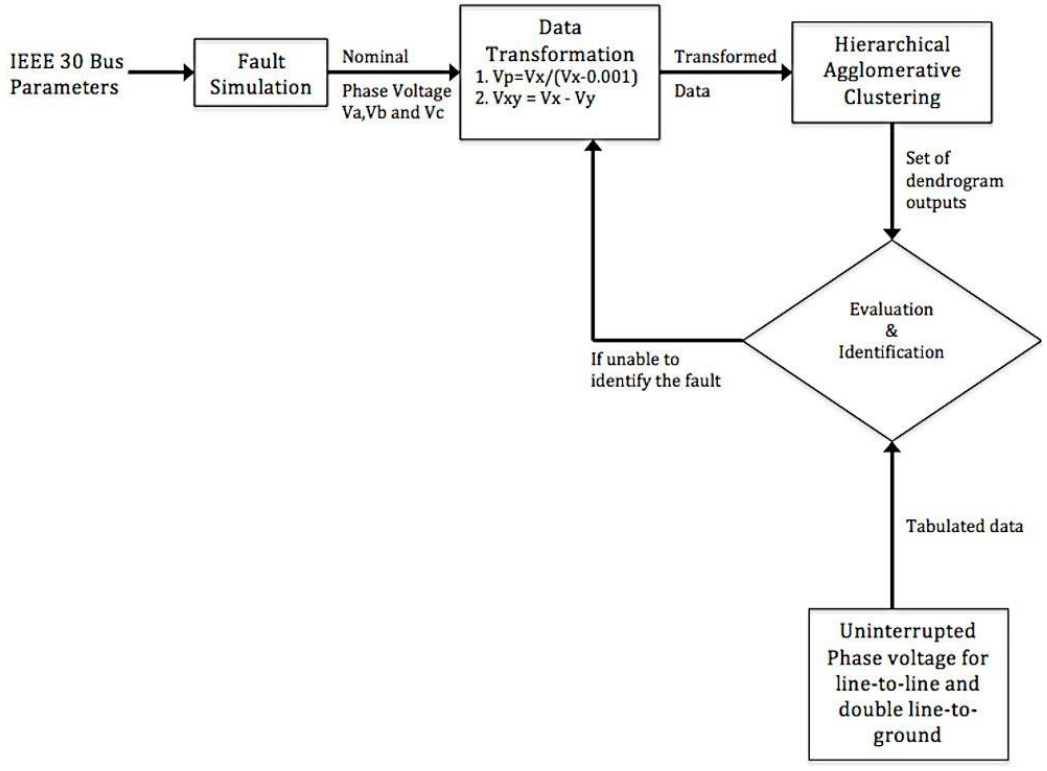

Fig. 3: Proposed cosine similarity cluster analysis model for transmission lines fault detection

3.1. Fault simulation of IEEE30 bus power system

The IEEE30 bus power system was simulated with Power World software v.18. The IEEE 30 bus power system had 6 generators, 30 buses and 41 numbers of transmission lines. The voltage magnitude of each bus was varied with a maximum of 1.05p.u and minimum of 0.95p.u. After all the parameters had been configured as shown in the Fig. 4, fault simulation was carried out by specifying types of faults and the bus at fault. Three phase nominal voltages at fault in each bus were produced by the simulation. These collected data were exported to SPSS software for further analyze the data using data mining technique.

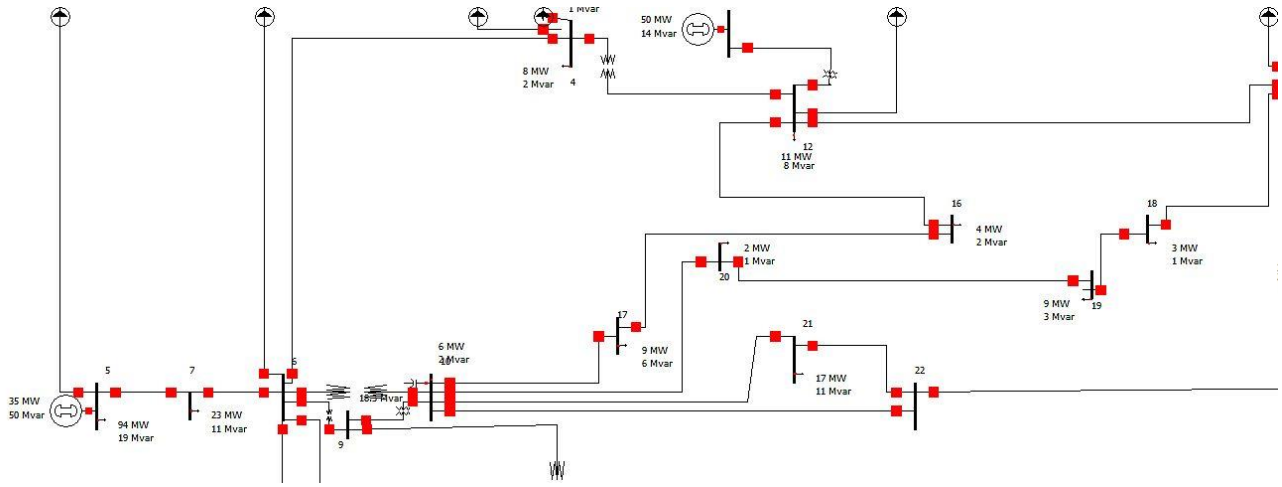

Fig. 4: One-line representation of IEEE30 bus power system

Three phase fault was simulated with Bus 2 at 3 phase fault (symmetrical fault) where all the phase voltages Va,Vb and Vc were 0 p.u volts. This was due to all the 3 phase were short circuited at once resulting high current flow and 0 potential differences between lines. Single line-to-ground fault was simulated with Bus 7 at single line-to-ground fault (asymmetrical fault) where only phase A voltage was short circuited to ground. This was because in Power World software by default the fault simulation for single line-to-ground fault was directed to phase A only. Thus, only phase A at Bus 7 was 0 p.u. Whereas, line-to-line fault was simulated with Bus 11 at line-to-line fault and double line-to-ground fault was simulated with Bus 18 at double line-to-ground fault. Both of the fault types were just affecting phase B and phase C lines, this was due to the Power World software default configuration of line-to-line fault and double line-to-ground fault.

3.2. Data transformation model

The performance of the Cluster Analysis was very dependent on the input data that was given to it. In order to enhance the performance of the clustering process in determining types of faults and the bus at fault, the raw data collected needed to be transformed into specific values. In this proposed model, there were two types of data transformation equations, each equation help to determine the specific type of faults and were executed in sequence. The first type of data transformation help to determine three phase fault, single line-to-ground fault and sometimes double line-to-ground fault, but in this paper, double line-to-ground fault was simulated in different conditions where the second process was required. The equation governing the first type of data transformation (Eqs. 4-6):

$$V'_{ai} = \frac{V_{ai}}{V_{ai}-0.0001} \quad (4)$$

$$V'_{bi} = \frac{V_{bi}}{V_{bi}-0.0001} \quad (5)$$

$$V'_{ci} = \frac{V_{ci}}{V_{ci}-0.0001} \quad (6)$$

where, V'_{ai}, V'_{bi}, and V'_{ci} are the result of preprocessed data and i is the bus number.

If with the first data transformation data at first process, the dendrogram unable to detect any anomaly bus at the output stage, then the raw data was subjected to the second type of data transformation at second process, the second type of data transformation help to determine line-to-line fault and double line-to-ground fault, the equation governing the second type of data transformation (Eqs.7-9):

$$V_{ab} = V_{ai} - V_{bi} \quad (7)$$

$$V_{ac} = V_{ai} - V_{ci} \quad (8)$$

$$V_{bc} = V_{bi} - V_{ci} \quad (9)$$

Cosine Similarity model of computing the proximity matrix performs precisely and fast with the transformed data. The data that had been transformed ease the process of clustering by reducing total number clusters and thus increasing the speed.

3.3. Hierarchical clustering transmission lines fault identification

The transformed data based on different type of fault were used as input for Cluster Analysis. The algorithm used for hierarchical clustering is Agglomerative based which was a bottom-up approach. Agglomerative clustering algorithm was used for the proposed model due to agglomerative fast and eases the clustering process with desired accuracy of the output. The agglomerative algorithm was shown below:

1. Compute the proximity matrix of the transformed phase voltages of each bus using Cosine Similarity equation below:

$$\frac{(V_{ai} \times V_{aj}) + (V_{bi} \times V_{bj}) + (V_{ci} \times V_{cj})}{\sqrt{V_{ai}^2 + V_{bi}^2 + V_{ci}^2} \times \sqrt{V_{aj}^2 + V_{bj}^2 + V_{cj}^2}}$$

where, i and j are bus number V_a, V_b, V_c are phase voltages of each bus.

2. Let each bus point be an individual cluster.
3. Each cluster was compared and similarity of each cluster was computed.
4. After similarity of each cluster was computed, merge two clusters which had the highest similarities to form a new cluster.
5. The proximity matrix was then updated as in step (1)
6. The algorithm iterates step (1) to step (5) until a single cluster was obtained.

3.4. Cosine similarity cluster analysis based data mining model flow chart

The overall model system flow chart was shown in the Fig. 5 below (Nithiyananthan et al., 2004). This flow chart will provide a clear idea on the overall process of the proposed model. Notice that there are two blocks of data transformations and repetition block of clustering and result evaluations. The 3 phase fault and single line-to-ground (SLG) fault can be determined during the first process.

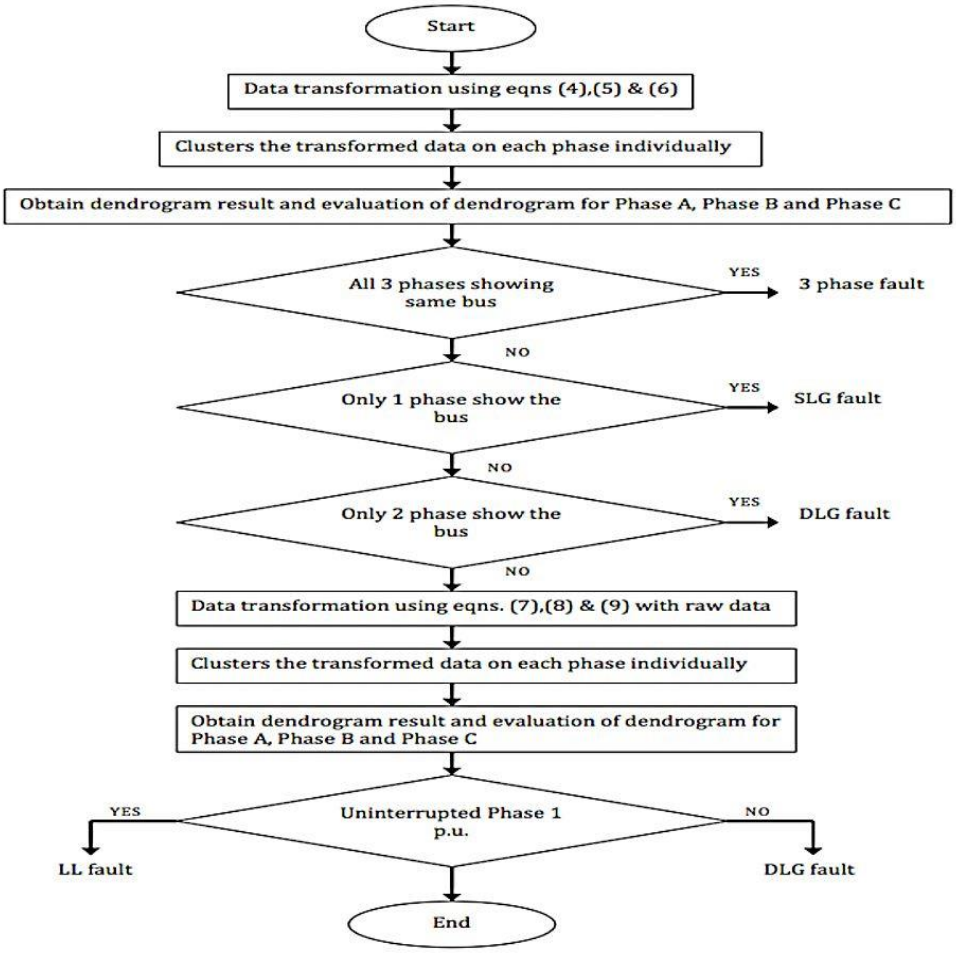

Fig. 5: Proposed system flow chart

Besides that, the third conditional block in the flow chart helps to determine double line-to-ground (DLG) too because some cases during the first process itself DLG fault was able to be determined. However, in this paper different case of DLG fault data was simulated which required for the second process to be executed in order to determine DLG fault. Line-to-line (LL) fault required the second process to be determined.

4. Results

The proposed system was simulated using Window 7 64-bits with SPSS software version 16.0 and PowerWorld software version 18. Cosine Similarity Cluster Analysis model performs very well together with suitable transformed data. Subsequent figures are the dendrogram result obtained from SPSS after Cluster Analysis was applied (Yong et al., 2015).

4.1 Phase fault identification

Fig. 6 shows that the proposed model successfully identified Bus 2 was at 3 phase fault as simulated in the Power World software. As shown in the above dendrogram Bus2 was the most dissimilar (rescaled distance) comparison to other bus which means there were anomaly occurred at the particular bus.

The dendrogram result of each phase was similar to Fig. 6, hence all phases had anomaly at Bus2. Thus, it can be concluded that Bus 2 was at 3 phase fault.

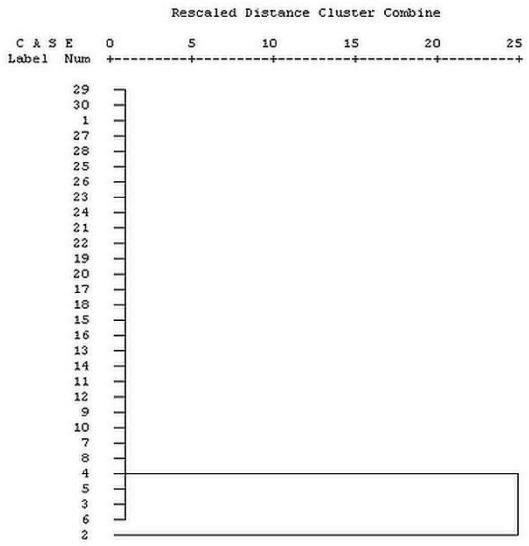

Fig. 6: Dendrogram of 3 phase fault identification

4.2. Single line-to-ground fault identification

Fig. 7 and Fig. 8 show the dendrogram results of single line-to-ground fault at different phases. As shown in Fig. 7, Phase A was able to identify accurately that Bus 7 was at fault and only cluster analysis for Phase A able to identify the bus at fault compare to Phase B and C shown in Fig. 8. Thus, Bus 7 at single line-to-ground fault at Phase A.

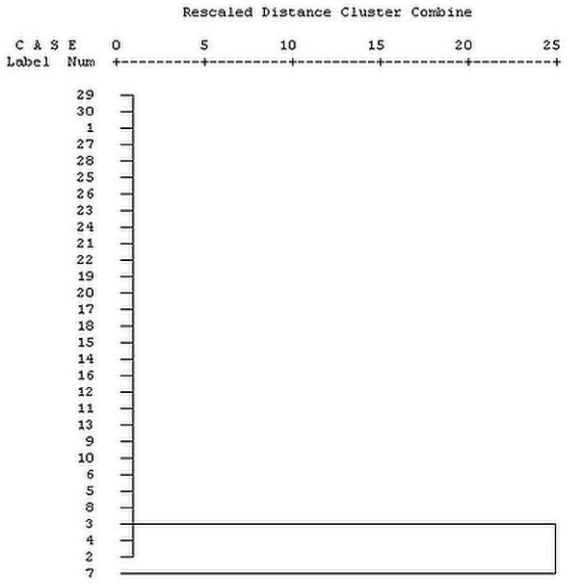

Fig. 7: Dendrogram of Phase A for single line-to-ground fault with data transformation (4), (5) and (6)

4.3. Line-to-line fault identification

The transformed data of line-to-line fault first analyze in the first process and there were no anomaly bus detected. Hence, the raw data collected from the fault simulation were subjected to data transformation in equation (7), (8) and (9) at the second process.

Fig. 9 shows that from the difference between Phase B voltages and Phase C voltages, Cluster Analysis identified Bus 11 was at fault.

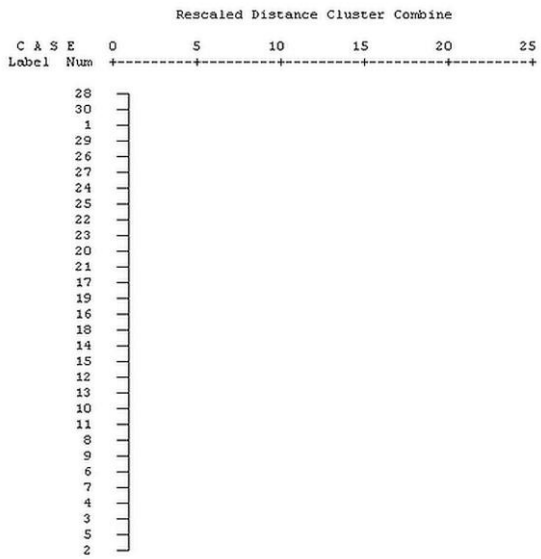

Fig. 8: Dendrogram of Phase B and Phase C for single line-to-ground fault with data transformation (4), (5) and (6)

Whereas, for Vab and Vac show that the difference between Phase A voltages and Phase B voltages and the difference between Phase A voltages and Phase C voltages there were no anomaly bus identified. Thus, the line Phase A was unaffected, in order to identify the Bus 11 was experiencing line-to-line fault the line Phase A voltages were at Bus 11 was checked. The Va phase voltage was equal or less than unity thus, Bus 11 was at line-to-line fault. If the Va phase voltages was more than unity, then that Bus will be at double line-to-ground fault.

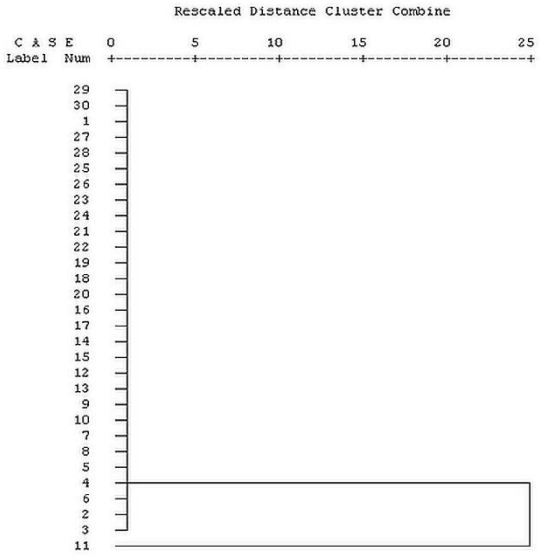

Fig. 9: Dendrogram of (Phase B-Phase C) for line-to line fault with data transformation (7), (8) and (9)

4.4. Double line-to-ground fault identification

In the first process there were anomaly buses detected in the cluster analysis, hence, the raw data were subjected to the second process.

Fig. 10 shows that from the difference between Phase B voltages and Phase C voltages, Cluster

Analysis identified Bus 18 was at fault with adequate accuracy. Whereas, Vab and Vac show that the difference between Phase A voltages and Phase B voltages and the difference between Phase A voltages and Phase C voltages there were no anomaly bus identified. Thus, the line Phase A was uninterrupted phase, in order to identify the Bus 18 was experiencing double line-to-ground fault the line Phase A voltages was at Bus 18 was checked. As can be observed that bus 18 Va p.u voltage was 2.18185 which more than unity, thus Bus 18 was experiencing double line-to-ground fault.

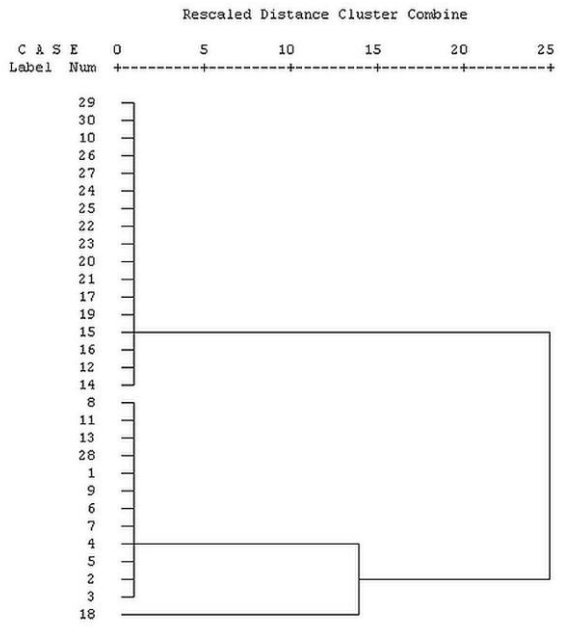

Fig. 10: Dendrogram of (Phase B-Phase C) for double line-to-ground fault with data transformation (7), (8) and (9).

5. Conclusion

An effective Custer Analysis based Power System model has been developed to identify symmetrical and asymmetrical transmission lines fault. The proposed model by using data mining technique such as Agglomerative clustering and Cosine model for calculation of proximity matrix had successfully identified 3 phase fault, SLG fault, LL fault and DLG fault in a 30 bus power system. This showed that there were huge potential by using DMT in power system, although very less research has been done. Therefore, in today's world of ICT there is a lot of information being presented, which can be utilized in order to optimize the efficiency in solving complex problems in power system. Conventional models will not perform as well as non-conventional models. DMT was one of the non-conventional models that are gaining popularity in analysis that involving large data in a database which having a huge potential in solving complex and practical problem.

Acknowledgement

We were very thankful for the Fundamental Research Grant Scheme (FRGS) funding from the Ministry of Higher Education (MOHE) Malaysia and AIMST University to carry out this research project.

References

Aldaz JM, Barza S, Fujii M, and Moslehian MS (2015). Advances in operator cauchy-schwarz inequalities and their reverses. Annals of Functional Analysis, 6(3): 275-295.

Dalstein T and Kulicke B (1995). Neural network approach to fault classification for high speed protective relaying. IEEE Transactions on Power Delivery, 10(2): 1002-1011.

Das B and Reddy JV (2005). Fuzzy-logic-based fault classification scheme for digital distance protection. IEEE Transactions on Power Delivery, 20(2): 609-616.

Ferrero A, Sangiovanni S, and Zappitelli E (1995). A fuzzy set approach to fault type identification in digital relaying. IEEE Transactions on Power Delivery, 10(1):169-175.

Jain AK, Murty MN, and Flynn PJ (1999). Data clustering: a review. ACM Computing Surveys (CSUR), 31(3): 264-323.

Mani P, Nithiyananthan K, and Nair P (2015). Energy saving hybrid solar lighting system model for small houses. World Applied Sciences Journal, 33(3): 460-465.

Nithiyananthan K and Ramachandran V (2002). EJB based component model for distributed load flow monitoring of multi-area power systems. International Journal for Engineering Modelling, 15(1): 63-67.

Nithiyananthan K, Ramachandran V, and Dhamodharan D (2004). RMI based multi-area power system load flow monitoring. Iranian Journal of Electrical and Computer Engineering, 3(1): 28-30.

Rohrig J (1931). Location of faulty places by measuring with cathode ray oscillographs. Elektrotech z, 8(2): 241-242.

Surya AV, Koley E, and Thoke AS (2014). Artificial neural network based fault locator for single line to ground fault in double circuit transmission line. International Proceedings of Economics Development and Research, IACSIT Press, Singapore 75: 47-51.

Tabatabaei A, Mosavi MR, and Rahmati A (2012). Fault location techniques in power system based on traveling wave using wavelet analysis and GPS timing. Electrical Review, 88(6): 347-350.

Takagi T, Yamakoshi YAMAURA, Yamaura M, Kondow R, and Matsushima T (1982). Development of a new type fault locator using the one-terminal voltage and current data. IEEE Transactions on Power Apparatus and Systems, 2(8): 2892-2898.

Wang H and Keerthipala WWL (1998). Fuzzy-neuro approach to fault classification for transmission line protection. IEEE Transactions on Power Delivery, 13(4): 1093-1104.

Yong T, Bin E, Nair P, Raguraman R, and Nithiyananthan K (2015). Local outlier factor based data mining model for three phase transmission lines faults identification. International Journal of Computer Applications, 130(2): 17-23.

Zhang Y, Jing MA, Zhang J, and Zengping WANG (2009). Applications of data mining theory in electrical engineering. Engineering, 1(03): 211-215.