Strategy for improvement the performance of a fault locator based on support vector machines

Estrategia para mejorar el desempeño de un localizador de fallas basado en máquinas de soporte vectorial

D. Arredondo[1], J. Mora-Florez[2], S. Pérez-Londono[3]

ABSTRACT

This paper presents a strategy aimed to improve the performance of a fault locator based on support vector machines (SVM). The strategy is based on the definition of the best number of operative conditions for parameterization and training stages, oriented to reach the better performance in minimal time intervals. The proposed methodology is tested in three distribution power systems designed according to the national technical standards for the different voltage levels. This paper shows the dependence relation between the training and parameterization databases with the performance of the fault locator

Keywords: Electric power distribution system, fault location, support vector machine, the training and parameterization stages.

RESUMEN

En este documento se presenta una estrategia implementada para mejorar el rendimiento de un localizador de fallas basado en la máquina de soporte vectorial (SVM). La estrategia se fundamenta en la definición del mejor número de condiciones de parametrización y entrenamiento para alcanzar buenos desempeños en un mínimo tiempo posible. La metodología propuesta se valida en tres circuitos de distribución diseñados con normas técnicas colombianas con diferentes niveles de tensión. Este análisis muestra que existen relaciones de dependencia entre la base de datos de entrenamiento y parametrización con el rendimiento del localizador.

Palabras clave: Sistemas de distribución de energía eléctrica, entrenamiento y parametrización, localización de fallas, máquinas de soporte vectorial.

Received: July 23th 2015
Accepted: Oct 15th 2015

Introduction

Service continuity in power distribution systems is a very important aspect for power utilities. These have to provide a good service, but faults at the power system are difficult to avoid. Faults affect the financial balance of the utility due to the cost of compensations for low continuity indexes, non-supplied energy and maintenance team workforce, among others [Dash, et al., 2007]

The most common fault in power distribution systems is the single-phase fault [Mora, 2006]. In most of the cases, these faults are difficult to localize due the length of the feeder, the level of the power distribution automation, presence of transient faults and the constrained availability of the maintenance staff. The restoration time strongly depends on the fault location time; nowadays in most of the utilities does not have any helping tool to facilitate the automatic fault location.

To solve this problem, several methodologies have been implemented and these methods use measurements of the voltage and current at the main power substation between fault and pre-fault stages [Agudelo, et al., 2014], [Orozco, et al., 2012], [Livani & Evrenosoglu, 2014]. There are two groups of fault location algorithms, the first is called methods based on the power system model and the second one is the method based on knowledge. The first one is based on the creation of a model of the power system to estimate the distance to the fault; this methodology has the problem of multiple estimation [Thukaram, et al., 2014] [Morales, et al., 2010]. The second one is based in data mining methods and has been used to determine the faulted zone and also to solve the multiple estimation problem [Gil, et al., 2013].

Data mining are techniques to extract information from the database. The size of the database is determinant in the perfor-

[1] Debbie Johan Arredondo Arteaga. B.Sc. in Electrical Engineering from the Universidad Tecnológica de Pereira and M.Sc candidate in Electrical Engineering from the Universidad Tecnológica de Pereira, Colombia. Email: djarredondo@utp.edu.co

[2] Juan José Mora-Flórez: B.Sc. and M.Sc. in Electrical Engineering from the Universidad Industrial de Santander, Colombia. Ph.D. in Electrical Engineering from the Universitat de Girona, España. Affiliation: Associated professor at Universidad Tecnológica de Pereira, Colombia. E-mail: jjmora@utp.edu.co

[3] Sandra Pérez-Londoño: B.Sc. and M.Sc. in Electrical Engineering from the Universidad Tecnológica de Pereira, Colombia. Ph.D. in Electrical Engineering from the Universidad Nacional de Colombia. Affiliation: Associated professor at Universidad Tecnológica de Pereira, Colombia. E-mail: saperez@utp.edu.co

VIII Simposio Internacional sobre Calidad de la Energía Eléctrica 2015 SICEL
Noviembre 18, 19 y 20 Valparaíso ■ Chile

mance of the fault locator, one of the most critical aspect in these methods is the training time, when there are large amounts of data, the time intervals and performance will be high but Little data often will reflect in minimal time intervals, and the performance will also be low. The actual estimation to determine the optimal size of the database is not an easy problem.

In this paper is proposed a strategy to improve the performance of the fault locator based on support vector machine (SVM). The main idea is to obtain the number of operative conditions (database of faults for an operative condition of the power system) for training and parameterization stages that reach the best performance of the locator. This methodology has been implemented for three power distribution systems designed with the Colombian standards. The support vector machine based fault locator belongs to the group of knowledge-based methods and as is described in several papers, the approach presents a good performance [Gil, et al., 2013], [Agudelo, et al., 2014]. However in most of the cases, the parameters are determined in a heuristic way, then number of operative conditions used in parameterization and training stages is here analyzed to determine the best relation between the computational effort and the performance of the fault locator.

This paper is divided in five sections. The section II presents the basic theoretical aspects, while the section III explains the proposed methodology. Section IV is devoted to show and analyze the obtained results, and finally, the section V presents the main conclusions obtained on the here reported research.

Theoretical aspects

This section is devoted to present the basic concepts used in this paper; detailed information are out of the scope of this section, however it can be obtained at the provided references.

A- Support vector machine

The support vector machine (SVM), is a classification technique that has recognized advantages over other classification techniques such as robust performance, generalization capability, single parameter dependency and the use of a kernel function, among others (references).

The SVM is a binary classification technique; however, it can be generalized to be a multiclass classification technique by using a decomposition-reconstruction technique has been reported in [Mora, 2006]. The here used decomposition and reconstruction strategies are next explained.

Decomposition one vs one (OVO): This strategy is used to convert the support vector machine from a binary classification machine into a multi-class machine. In this phase, a multi class problem is decomposed in N binary classification problems [Mora, 2006].

Reconstruction by simple vote: This phase is aimed to integrate the results of the N binary problems and emulate the behavior of a single multi classification machine. This voting scheme is used in the here presented research due to its easy to implementation and the reduced quantity of ties between classes [Gutiérrez, et al., 2010], [Morales, et al., 2010].

B- Data base generation

The fault database required to develop the proposed approach, is obtained using several tools created by the research group. Some of the most relevant used tools are briefly presented.

RF simulation is an algorithm developed to generate faults in a power distribution system, which is based in techniques as Latin hypercube, sensitivity analysis, among others, to create an amount of operative conditions by variations on loads, supply voltage, power frequency, among others. In those conditions, the algorithm performs faults in every node of the analyzed power system, and voltages and currents at the main power substation are measured and stored [Mora, et al., 2006]. This tool uses an efficient combination of the Alternative transients program (ATP) and MATLAB.

Latin hypercube is the here used sampling technique, because it is independent of the analyzed model. It also reduces the number of evaluations, generating a small dataset that fully represents the total sampling space [Alzate, et al., 2014]. This technique is used here to determine the operative conditions of the analyzed power distribution system, to be simulated wit RF simulation tool in order to obtain the fault database.

Proposed methodology

The figure 1 shows the proposed methodology scheme. This scheme consists of five different stages aimed to improve the fault locator performance.

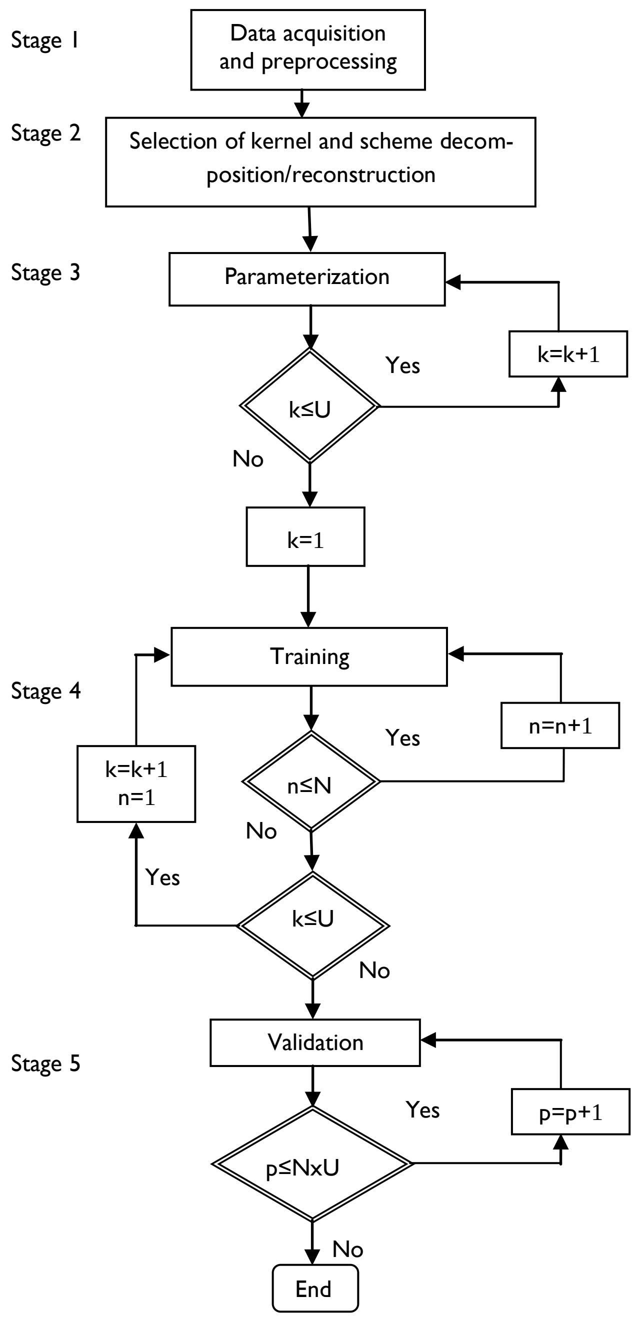

Figure 1. Methodology scheme. Source: authors

A- Stage 1: Data acquisition and preprocessing,

Data acquisition: Using the Latin hypercube technique, the variations from the rated conditions of the power distribution system are obtained. Then, by using the simulation RF tool, faults are performed in each node and using different values of fault resistance, at the analyzed power distribution system.

Data preprocessing: The data processing improves the performance of SVM while the computational effort is reduced. This process consists of feature extraction and normalization of database, which aims to eliminate unnecessary or inconsistent information for the SVM [Gil, et al., 2013]. At this stage, features as the magnitude and angles of voltages and current measured at the main power substation and during pre-fault and fault conditions, are obtained.

B- Stage 2: Selection of kernel and scheme decomposition/reconstruction

Type of kernel: Several kernels can be used, however the radial basis function (RBF) was selected due the adequate performance in all the fault location researches [Agudelo, et al., 2014], [Thukaram, et al., 2014], [Mora, 2006], [Gutiérrez, et al., 2010].

Scheme of decomposition and reconstruction: At the proposed methodology the one vs one (OVO) decomposition method has been used. In addition, single voting reconstruction method is also used. These methods transform the binary classifier in a multi-class classifier as previously explained, [Gil, et al., 2013].

C- Stage 3: Parameterization

At this stage, the optimal parameters of the classifier are determined (C and γ, where C is the penalty parameter for SVM and γ is the kernel parameter). The parameterization stage in the proposed method is an iterative phase due to the use of a determined number of operative conditions. k at figure 1, it is a counter of the number parameterizations performed; this process is repeated U times.

The Tabu metaheuristic searches the optimal parameters using as objective function the cross-validation error. Then the algorithm obtained the smallest cross-validation error, as presented in [Gil, et al., 2013].

Parameterization: This phase is the important due to the optimization algorithm (metaheuristic Tabu) because, this technique improve the separation of data. The rated operative condition is needed for this process because it is recommend using less than 1% of the all operative conditions at the database (the parameterization must to be a fast process).

The parameterization in the scheme (figure 1) must to be done with different numbers of conditions; the final idea is to determine an optimal number of parameterization conditions that improve the performance of the locator.

In this paper has been used one, three, five, seven and nine operative conditions, including the nominal condition, for the parameterization process, then U=5.

D- Stage 4: Training

At this stage, optimal parameters obtained for every parameterization are used (U parameterization). According to the previously exposed, the machine must to be trained several times (NxU).

The training process is computationally expensive; however the number of training conditions is determinant in the final performance. Therefore it is recommended use less than 10 % of the totals power system operating conditions.

The training process is performed for different numbers of conditions (N). Moreover it is important that all the training conditions have to be done for every single parameterization (k).

In training process, n and k are training counters. The numbers of training conditions chosen are 50, 100, 150, 200, 250, 300, 350, 400 and 700, then N=9. This number of conditions was choosing because the performance is good and the computational cost is not very expensive.

E- Stage 5: Validation

Using the fault database, each training process in stage 4 is evaluated for the remaining of the database not used in parameterization and training stages.

The validation phase is performed for all of the training conditions analyzed on stage 4. This process is done with more than 80 % of the total database, in this phase is determined the performace of the learning technique.

The parameter p in the scheme (figure 1), it is a counter that increments with each validation set used at the training process. The total number of training conditions is NxU, (N total number of training for each parameterization conditions, U total number of parameterization conditions).

Obtained results

At this research three distribution power systems was selected to validate the proposed strategy; these power systems are a good replica of the currently installed power distribution systems in Colombia, and probably in several regions of the world [Livani & Evrenosoglu, 2014].

The tests circuits are shown in figures 2, 3 and 4, for these circuits was obtained a large database with 5000 operative conditions.

Urban power networks, at 13.2kV rated voltage: these power distribution systems are common in the urban zones, which are mainly characterized by the presence of many single and three phase laterals and loads of relatively low power. A sample of such power systems is shown in figure 2.

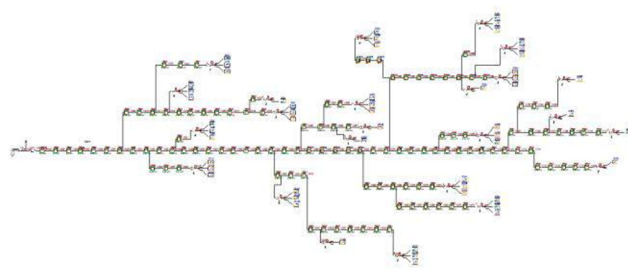

Figure 2. Urban Network at 13.2 kV rated voltage. Source: authors

Rural power networks, at 34.5 kV rated voltage: many of the power distribution systems in Colombian, have circuits, which are designed to operate at this voltage level. These are used to supply small towns and medium industries. A single line layout of this system is shown in figure 3.

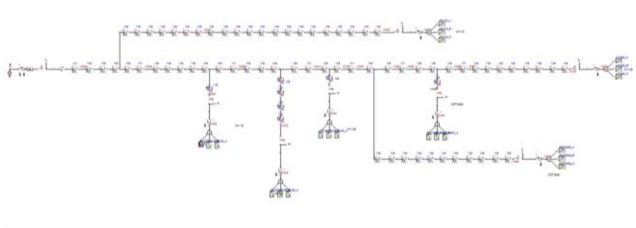

Figure 3. Rural Network at 34.5 kV rated voltage. Source: authors

Rural power network, at 44 kV rated voltage: this type of power systems can be classified as sub transmission lines [Epm, 2009]. A sample of this power feeder is shown in figure 4

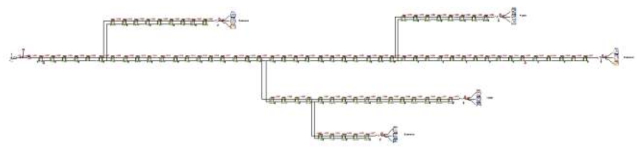

Figure 4. Rural Network at 44 kV rated voltage. Source: authors

In figures 5, 7, 9 show the performance of the fault locator related to the number of training and parameterization conditions for each analyzed power system. In addition, the training time for each circuit is illustrated in figure 6, 8 and 10. Furthermore these graphs are important because shows that better performance is computationally expensive.

Parameterization: Five different sets of operating conditions for each circuit were used in training. The different used conditions are: 1- rated condition, 2- rated condition plus two random obtained operating conditions, 3- rated condition plus four random obtained operating conditions, 4- rated condition plus six random obtained operating conditions, 5- rated condition plus eight random obtained operating conditions.

Training: At this phase, 700 operating conditions were used for each circuit. Sets of 50, 100, 150, 200, 250, 300, 350, 400 and 700 random operative conditions were considered. More operating conditions were not tested because the not affordable computational cost.

Validation: in this phase 4000 operating conditions for each circuit were used. The training models obtained at stage 4 are evaluated for each operative condition.

A- Urban power network, at 13.2kV rated voltages

The performance of the fault locator at the selected training conditions are shown in figure 5. The variable at the horizontal edge means the training conditions. The variable at the vertical edge mean, the performance of the locator. The performance is the test of the validation condition in each training model (stage 4).

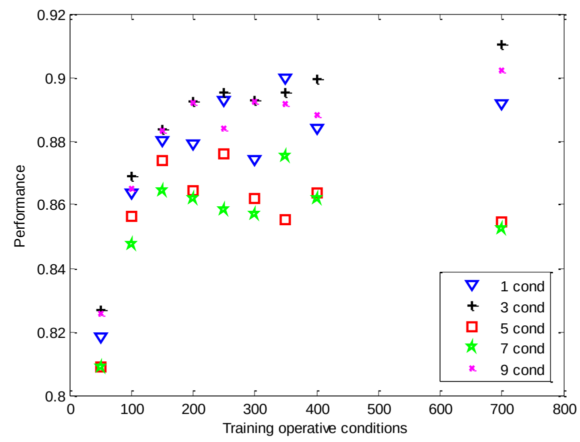

Figure 5. Performance of the fault locator versus the number of operating conditions used in training. Source: authors

As shown in figure 5, the performance increases as the number of training and parameterization conditions does. The training using more than 200 operative conditions increases the performance for some parameterization conditions, but there is not a big change. On the other hand, as is shown in figure 6, the time for training has an exponential behavior, then the use of more than 200 conditions are not recommended.

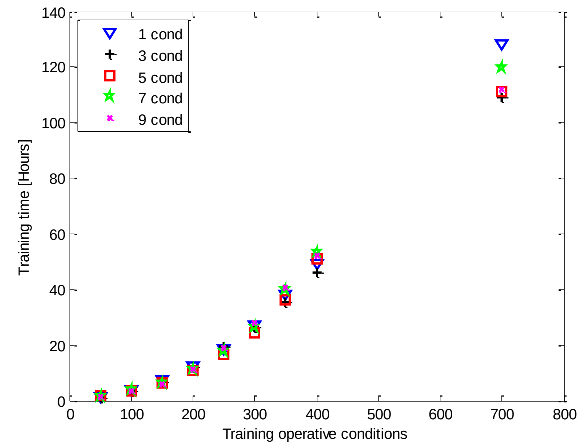

Figure 6. Time training behavior versus the number of operating conditions used in training. Source: authors

Figure 6 shows that the different sets of parameterization conditions have an insignificant increment at the training time. However the computational time is very sensible to the number of training conditions, then 200 operative conditions it is a good number to choose due to the demonstrated performance of the locator and the computational effort at the training time.

B-Rural power network, at 34.5 kV rated voltage

The performance of the locator of faults is illustrated in figure 7 for parameterization and training conditions

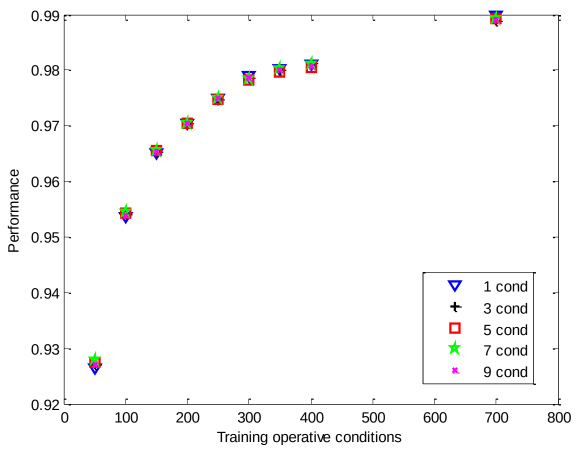

Figure 7. Performance of the fault locator versus the number of operating conditions used in training. Source: authors

Figure 7 show that the parameterization conditions do not affect the performance. Secondly, the number of training conditions has dependency with performance but 300 training conditions is a good candidate.

In figure 8 are shown the time training conditions.

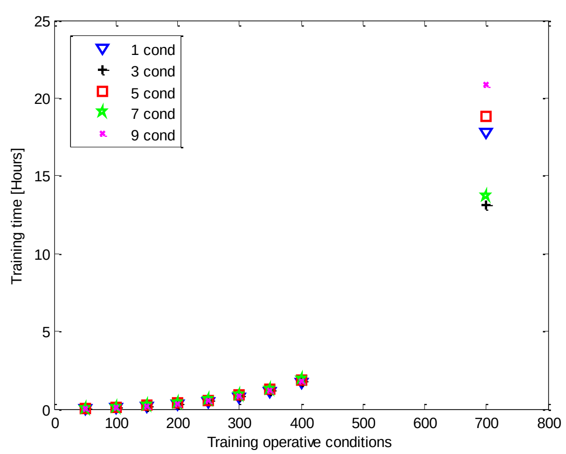

Figure 8. Time training behavior versus the number of operating conditions used in training. Source: authors

The parameterization conditions do not affect the training time however the training conditions is determinant, and it is proposed choose 300 conditions due to the low computational time cost and the performance reached with this conditions

C- Rural power network, at 44 kV rated voltage

In figure 9 are shown that the better performance is reached by 7 conditions of parameterization and 300 training operative conditions.

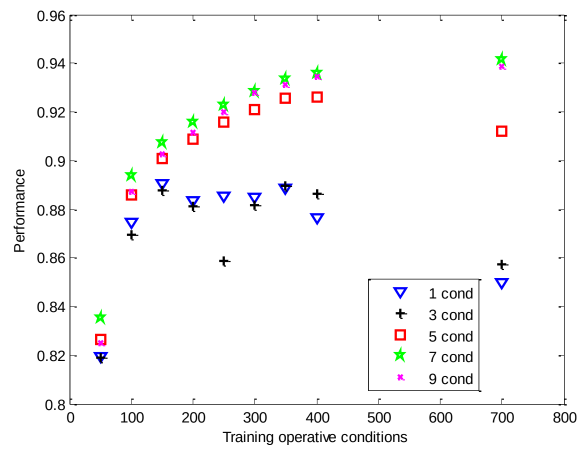

Figure 9. Performance of the fault locator versus the number of operating conditions used in training. Source: authors

The figure 9 shows that the locator is sensible to the parameterization conditions like figure 6 and the figure 10 shows the training computational efforts for 44 kV circuit.

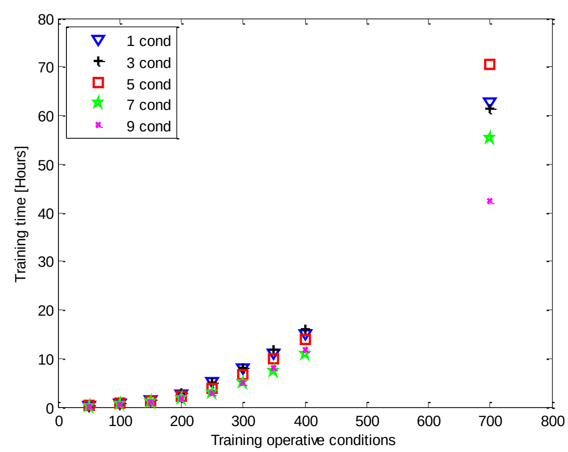

Figure 10. Time training behavior versus the number of operating conditions used in training. Source: authors

In the figure 10, it is validated the supposed in figure 9; the best performance is reached by 300 operative training conditions and seven parameterization conditions. As result, the locator of fault reached a better efficiency in an optimal training time.

In the table I are summarized the results for all of the analyzed power distribution systems.

Tablet I Summary result. Source: authors

13.2 kV	
Parameterization condition	Training condition
3	200
34.5 kV	
Parameterization condition	Training condition
I	300
44 kV	
Parameterization condition	Training condition
7	300

Conclusions

The strategy proposed allows choosing an optimal number of parameterization and training conditions, and shows the effectiveness of the fault locator based on machine learning. On the other hand, a general method for a fault location based on support vector machines is not an easy problem, due the need of selection of the adequate parameters to assure good performance.

For each analyzed power system, the method here proposed helps to obtain the operative conditions that improve the performance of the fault locator.

Finally, the setting helps to improve the performance of the SVM based fault locator, providing a useful tool to improve the power continuity at power distribution utilities.

7. Acknowledgments

This research was supported by the Universidad Tecnológica de Pereira (Colombia) and COLCIENCIAS, under the project "Desarrollo de localizadores robustos de fallas paralelas de baja impedancia para sistemas de distribución de energía eléctrica LOFADIS 2012", contract number 0977-2012. and the Master program in Electrical Engineering Universidad Tecnológica de Pereira.

References

Agudelo, C., J.; Gil, G., W.; Escobar, Z., A.; Mora F., J., Analysis of the effects of load model on a single phase fault locator based on support vector machines., Transmission & Distribution Conference and Exposition - Latin America (PES T&D-LA), 2014 IEEE PES , vol., no., 10-13 Sept. 2014, pp.1-5

Alzate, G., N.; Mora, F., J.; Perez, L., S., Methodology and software for sensitivity analysis of fault locators., Transmission & Distribution Conference and Exposition - Latin America (PES T&D-LA), 2014 IEEE PES , vol., no., 10-13 Sept. 2014, pp.1-6

Dash P. K., Samantaray S. R., Panda G., "Fault classification and section identification of an advanced series compensated transmission line using support vector machine.," IEEE Trans. Power Del., vol. 22, no. 1, Jan. 2007, pp. 67–73.

EPM. Public Enterprises of Medellín, "Standards and Technical Specifications" (Normas y especificaciones técincas), 2009. [Online]. Available:https://www.epm.com.co/site/Home/Centrodedocumentos/Pro veedoresycontratistas/NormasyEspecificaciones/Normasa%C3%A9reas/ Grupo2Normasde132kV.aspx. [Last access: July 2013]

Gil W., Mora J., Pérez S., Comparative analysis of metaheuristics optimization techniques to parameterize fault locators for power distribution systems., Ingeniería y Competitividad, vol. 15, no. 1, 2013, pp. 103-115.

Gutiérrez J., Mora J., Pérez S., Strategy based on genetic algorithms for an optimal adjust of a support vector machine used for locating faults in power distribution systems., Revista de la Facultad de Ingeniería, no. 53, Jun. 2010, pp. 174-187.

Livani, H., Evrenosoglu, C.Y., A Machine Learning and Wavelet-Based Fault Location Method for Hybrid Transmission Lines., Smart Grid, IEEE Transactions on , vol.5, no.1, Jan. 2014, pp.51-59.

Mora J., Localización de faltas en sistemas de distribución de energía eléctrica usando métodos basados en el modelo y métodos de clasificación basados en el conocimiento, Ph.D thesis, Universidad de Girona, España, 2006.

Mora, J.,J.; Bedoya, J.,C.; Melendez, J., Extensive Events Database Development using ATP and Matlab to Fault Location in Power Distribution Systems., Transmission & Distribution Conference and Exposition: Latin America, 2006. TDC '06. IEEE/PES , vol., no., 15-18, Aug. 2006, pp.1-6.

Morales G., Mora J., and Vargas H., Fault location method based on the determination of the minimum fault reactance for uncertainty loaded and unbalanced power distribution systems., Transmission and Distribution Conference and Exposition: Latin America (T&D-LA), 2010 IEEE/PES, Sao Paulo, Brazil, 8-10, Nov. 2010

Orozco, H.,C., Mora, F, J., Perez L, S., A robust method for single phase fault location considering distributed generation and current compensation., Transmission and Distribution: Latin America Conference and Exposition (T&D-LA), 2012 Sixth IEEE/PES, 2012, pp.1-7.

Osorio, P, J.; Mora, F, J.; Rios, P, C., Characterization of the power distribution systems as test bed for fault locators., Transmission & Distribution Conference and Exposition - Latin America (PES T&D-LA), 2014 IEEE PES , vol., no.,10-13 Sept. 2014, pp.1-6

Pérez, L., Mora, J., Pérez, S., Design of an efficient and automated fault simulation tool applied to electric power systems., Dyna, vol. 77 no. 164, Dec. 2010, pp. 178-188.

Salat, R., Osowski, S., Accurate fault location in the power transmission line using support vector machine approach, IEEE Trans. Power Syst., vol. 29, no. 2, May 2004, pp. 979–986.

Salim, R.,H., Salim, K.,C.,O., Bretas, A.,S., Further improvements on impedance-based fault location for power distribution systems, IET Generation, Transmission and Distribution, Vol. 5, 2011, pp. 467–478.

Thukaram, D., Khincha, H., Vijaynarasimha, H., Artificial Neural Network and Support Vector Machine Approach for Locating Faults in Radial Distribution Systems., IEEE Transactions on Power Delivery, vol. 20, no. 2, Apr. 2005, pp. 710-721.