Universidad Industrial de Santander

Comparison of measurement features used as inputs in a learning-based fault location method for power distribution systems

Comparación de los descriptores utilizados como entradas en un método de localización de fallas basado en el aprendizaje para sistemas de distribución de energía

Ever Correa-Tapasco[1a], Juan Mora-Flórez[1b], Sandra M. Pérez-Londoño[1c]

[1]Grupo de Investigación en Calidad de Energía Eléctrica y Estabilidad - ICE3, Electrical Engineering Program, Universidad Tecnológica de Pereira, Pereira, Colombia. Email: [a] ever@utp.edu.co, [b] jjmora@utp.edu.co, [c] saperez@utp.edu.co

Abstract

This paper presents a comparative study of the measurement features used as inputs of a fault locator based on Support Vector Machines, which is aimed to analyze single-phase faults. Studies have shown that a huge database is required to obtain high performance, but a problem is associated with the excessive computing time required to evaluate such databases. This study examines properly these inputs to determine which are the most significant ones in terms of performance. Tests are performed on a 75 bus 34.5 kV distribution system, with 75000 shunt faults, implemented in ATP. According to the results, 12 features related to magnitude variations of phase voltage and current between fault and pre-fault steady states were relevant to achieve a performance of 96.3%, with a computational time of training and cross-validation of approximately six minutes.

Keywords: attributes; distribution systems; fault location; measurement features; support vector machines.

Resumen

En este artículo se presenta un estudio comparativo de descriptores utilizados como entradas en un localizador de fallas, basado en máquinas de soporte vectorial, cuyo objetivo es analizar fallas monofásicas. Estudios han demostrado que, para obtener un alto rendimiento, se requiere una gran base de datos, pero existe un problema que está asociado con el excesivo tiempo de cómputo necesario para analizar esas bases de datos. Este estudio contribuye a la solución del problema, pues analiza adecuadamente estas entradas del método y descubre cuáles son las más significativas. Las pruebas se realizan en un sistema de distribución de 34,5 kV, 75 nodos con 75000 fallas, implementado en ATP. De acuerdo con los resultados, doce descriptores relacionados con variaciones en magnitud de la corriente y la tensión de fase entre estados de falla y prefalla fueron relevantes al lograr un desempeño de 96,3 %, con un tiempo computacional de entrenamiento y validación cruzada de aproximadamente seis minutos.

Palabras clave: atributos; descriptores; localización de fallas; máquinas de soporte vectorial; sistemas de distribución.

1. Introduction

Power quality is one of the main issues in power distribution systems. The service discontinuity is a problem, where the design and the fault management play an important role as alternative solutions in distribution systems. Unfortunately, due to their stochastic nature, distribution systems faults are hardly avoidable [1], [2], [3].

Fault locators help to reduce the problem. First, a fault location helps to speed up the restoration process. Second, by locating the fault it is possible to perform switching operations to reduce the affected area. Moreover, the location of non-permanent faults allows scheduling maintenance tasks to avoid future problems [4], [5], [6]. Some fault location methods rely on impedance calculation estimated from the substation. The main disadvantage is associated with the multiple estimations of the fault location and the high dependency of the model [6], [7].

On the other hand, many researchers have recently addressed the problem of using Learning-Based Methods (LBM), whose objective is to exploit the existence of previous experiences and contextual information [2], [4], [8]. Thereby; it is possible to eliminate the multiple estimations. One of the learning algorithms for data analysis is the Support Vector Machines (SVM), which is based on statistical learning theory, quadratic programming and several clearly defined constraints as well as kernel transformations. SVM is fed with a database that includes voltage and current measurements at a single end and the topology of the power distribution system. SVM is known to have good results in diagnosis applications, especially to determine the faulted zone. Therefore, LBM in this work resorts to SVM as a classification technique (SVM-c) [2], [8], [9].

However, SVM has a high computational effort because it is required to process a large amount of data to adequately represent the problem and achieve satisfactory performance results. Studies have focused on computational efficiency, and different strategies have been used, namely, the implementation of several training methods with the comparison of results [2], parameters for the calibration of the learning method [10], database normalization [11] and analysis of the method inputs to discover which ones are the most

relevant [12], [13], [14]. This work is a follow-up to [12], where measurement features, including variables related to voltage and current, had the most significant contribution to fault location [15], [16]. The measurement features of this work will not consider variables such as reactance, apparent power and power factor. The focus herein will be exclusively on phasor voltages and currents.

The voltage and current measurements are collected from an IED located in the distribution system main substation. The relay takes a predetermined number of samples in pre-fault and other in fault situation, to get both the phase voltage and the line current. The appropriate sampling frequency of the relays to achieve the fundamental of the 60 Hz signal can be 16 or 32 samples per cycle. Then, with this data the phasors of pre-fault and fault steady states are estimated. Finally, from the phasor data, the measurement features used in this paper, are extracted.

The input of methods should be the result of the selection of measurement features and the identification of the most significant characteristics of a database for SVM-c. This is to achieve the most optimal predictive performance of the classifier with the minimum effort [12], [13], [14].

2. Methodological approach

The methodology is divided into three sections represented in the general scheme of Figure 1. Firstly, there is a brief description of SVM-c linear and nonlinear cases (stages 3 and 4). The second section defines the parameters necessary to apply LBM in fault location and shows how the LBM performance is calculated (stage 2). Finally, the third section specifies the general form of the measurement features used in this study (Stage 1) [13].
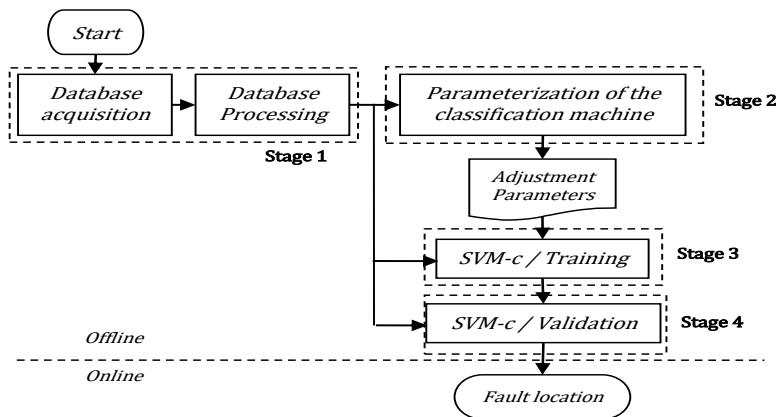
Figure 1. General scheme of the classification method. Source. [13].

Comparison of measurement features used as inputs in a learning-based fault location method for power distribution systems

75

2.1. LBM basics using SVM-c

2.1.1. LBM linear case

Classification using SVM involves training and testing data, which is composed of many occurrences summarized in (1). In training set, each occurrence consists of x_i attributes (measurement features) in an N-dimensional space and a target value (y_i) called class label (usually 1 or -1; two fault zones) [5], [7].

$$x_i \in R^N \ and \ y_i \in \{+1, -1\} \qquad (1)$$

The aim of this classifier is to create a model, which can successfully predict the class label (fault zone) from x_i (measurement features).

2.1.2. LBM non-linear case

As in the case of non-linear separable feature sets (zones), it is possible to transform the input into a new higher dimension space, where the data (zones) are linearly separable. A transformation function *F(.)* is defined in terms of inner products of the input data in the original classification space; such transformation is achieved in a single step by applying the corresponding kernel function for each case. Thus, linear classification algorithms can be extended to non-linear cases [5], [7]. When a Radial Basis Function (RBF) is chosen as kernel function, two SVM penalization parameters (constant *C* and kernel parameter **σ**) must be fixed by means of grid search and cross-validation in order to regulate the allowance of errors in databases [4], [10]. In this paper, the Gaussian RBF kernel is used and presented in Equation (2).

$$k(x,y) = exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right) \qquad (2)$$

In the representation of non-linear cases in Block diagram of Figure 2, databases and parameters C and σ can be seen as the two inputs of SVM-c. The database includes variable parameters (measurement features) and several fixed parameters of the power distribution system (zones, fault scenarios, number of fault records and normalization) [18].

2.2. SVM-c as a learning-based fault location method

The design process of LBM starts with preparation of a suitable training data set comprising of all possible fault scenarios that SVM-c needs to learn. Previously, nodes of a selected power system must be classified and grouped in zones, according to recommendations of the grid operator.

Different operating scenarios of the system are considered, the respective fault registers are obtained in each node of the system, the measurement features are extracted, and a process of normalization of the data is carried out. Normalization limits the values from the database within a range, usually between zero and one, which can also improve the accuracy, efficiency and computational times of the SVM [11].

The validation step of LBM is repeated n times by using a different subset (cross-validation) and consequently, a different combination of validation subsets. The performance is expressed by the ratio of the number of faults correctly located and the total number of faults, as seen in Equation (3) [5].

$$Performance = \\ = \frac{Number \ of \ faults \ well \ classified}{Total \ number \ of \ faults} \text{x}100\ \% \qquad (3)$$
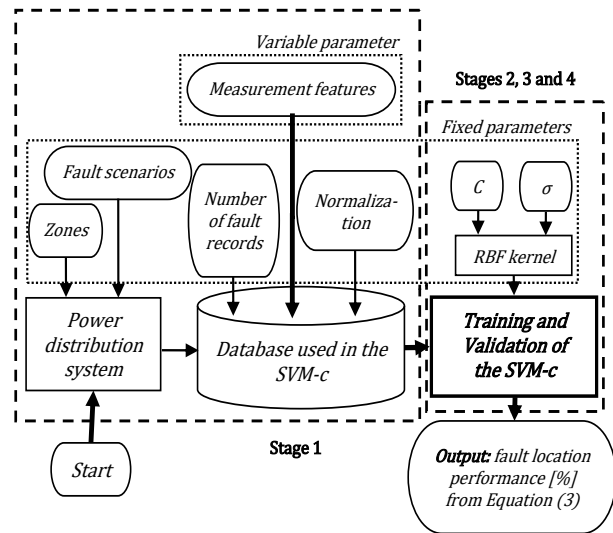

Figure 2. Block diagram of the implementation of fault locator method for non-linear cases. Source. Own elaboration.

2.3. Measurement features used as inputs in the learning-based fault location method

The assessment of SVM performance is based on the comparison of capacity of measurement features to contribute on fault location. Table 1 presents the nomenclature corresponding to the measurement features considered in this study. Variations are regarded as

difference of a variable between fault and pre-fault steady states.

Only phasor measurements of phase voltage and line current are available. Additional data are gathered through the corresponding linear combinations for each case. For example, to attain line voltage data, the corresponding linear combination is used with the phase data.

3. Results

Tests are performed on the power distribution system 34.5kV-75 bus test feeder implemented in ATP, and nodes are classified into five zones as recommended by the operator (Figure 3). This test feeder is a prototype distribution network, which represents a rural circuit of the Colombian primary distribution that connects a large number of towns with small urban and extensive rural areas. The circuit was developed to validate fault location methods under conditions close to reality.
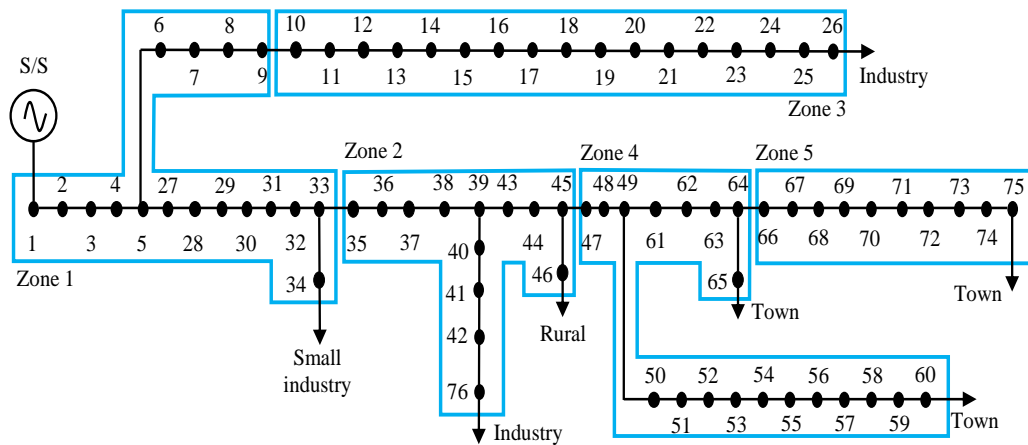

Figure 3. System 34.5kV-75 bus test feeder implemented in ATP. Source. Own elaboration.

LBM has been implemented in MATLAB environment using Lib-SVM toolbox [17], [19]. The fault database is obtained from a collaborative strategy between ATP and MATLAB [20]. It contains 75.000 records of single-phase faults derived from five values of fault resistance (0.0002 Ω, 10 Ω, 20 Ω, 30 Ω, and 40 Ω), 200 operating conditions of the system and 75 nodes located along the main feeder of the test feeder (from 2 to 76) [21]. Table 2 summarizes the conditions under which the tests are performed on the system.

The SVM-c requires a representative fault database of the distribution system. In this paper, 200 operating conditions are used. Each operating condition represents system scenarios where variations of system parameters such as load, signal frequency and/or magnitude of voltage are considered. However, it is possible that a specific distribution feeder requires, more or less, conditions to represent adequately the fault analysis.

Table 3 discriminates between vectors cases involving the three phases and those that compile data exclusively from the faulty phase. Vector cases 1, 7, 9 and 10 belong to the first group. From them, only faulty-phase vector cases 4, 8, 11 and 12 are derived respectively. To clarify this classification, features in the table suffixed with *(a)* denote one single value for the faulty phase, and features without the suffix are related to three values, one value for each phase. Cases 2 and 3 separate in-phase data from in-line data included in Case 1. Likewise, cases 5 and 6 are derived from case 4.

Table 4 shows the performance, computational time and the number of measurement features for each case shown in Table 3.

Comparison of measurement features used as inputs in a learning-based fault location method for power distribution systems

77

Table 1: Measurement features of system

Abbreviation	Feature description
dI	Variation in phase current magnitude.
dI_L	Variation in line current magnitude.
dV	Variation in phase voltage magnitude.
dV_L	Variation in line voltage magnitude.
$dV_L(a)$	Variation in line voltage magnitude, phase: a.
$dAngI$	Variation in magnitude of the phase current angle.
$dAngI_L$	Variation in magnitude of the line current angle.
$dAngV$	Variation in magnitude of the phase voltage angle.
$dAngV_L(a)$	Variation in magnitude of the line voltage angle, phase: a.
I_p	Phase current magnitude in pre-fault steady state.
I_{Lf}	Line current magnitude in fault steady state.
V_p	Phase voltage magnitude in pre-fault steady state.
V_f	Phase voltage magnitude in fault steady state.
$V_f(a)$	Phase voltage magnitude in fault steady state, phase: a.
$AngI_f$	Phase current angle in fault steady state.
$AngV_f(a)$	Phase voltage angle in fault steady state, phase: a.
$|dI|$	Phasor variation magnitude of the phase current.
$Ang(dV)$	Phasor variation angle of the phase voltage.
$|dIa|$	Phasor variation magnitude of the phase current, phase: a.

Source. Own elaboration.

Table 2. Test conditions

Parameter	Value
Kernel	Radial base function (RBF)
Normalization	Min-Max [11]
System state	Different conditions
Zone	Default
Kernel penalty parameters	C=512 σ =0.261215
Faults number	75000
Subsets of cross validation	5
Faults number to be validated	15000
Measurement features	**Table 3**

Source. Own elaboration.

According to Table 4, the measurement features for cases 1 and 2 are useful to achieve the best performances. A performance of 97.2% with measurement features 1 is obtained, and it indicates that in the case of 1000 faults, 972 of these faults are properly estimated, that is, they are within the zone in which it was designated as faulted zone. Likewise, case 1 shows a time of approximately 8 minutes. Which is a bit greater compared to case 2 due to the difference in the number of used measurement features.

Table 3. Combinations of measurement features.

Cases of Measurement Features	Measurement Features (attributes: x_i)
1	$dI, dI_L, dV, dV_L, dAngI, dAngI_L, dAngV, dAngV_L.$
2	$dI, dV, dAngI, dAngV.$
3	$dI_L, dV_L, dAngI_L, dAngV_L.$
4	$dI(a), dI_L(a), dV(a), dV_L(a), dAngI(a), dAngI_L(a), dAngV(a), dAngV_L(a).$
5	$dI(a), dV(a), dAngI(a), dAngV(a).$
6	$dI_L(a), dV_L(a), dAngI_L(a), dAngV_L(a).$
7	$|dI|, |dV|, Ang(dI), Ang(dV).$
8	$|dI(a)|, |dV(a)|, Ang[dI(a)], Ang[dV(a)].$
9	$I_p, I_f, V_p, V_f, AngI_p, AngI_f, AngV_p, AngV_f.$
10	$I_f, V_f, AngI_f, AngV_f.$
11	$I_p(a), I_f(a), V_p(a), V_f(a), AngI_p(a), AngI_f(a), AngV_p(a), AngV_f(a), .$
12	$I_f(a), V_f(a), AngI_f(a), AngV_f(a).$

Source. Own elaboration.

Table 4. Results summary of the ten cases of measurement features.

Cases of Measurement Features	Performance [%]	Time [minutes]	Number of Measurement Features
1	97,2	8,013	24
2	96,3	6,026	12
3	94,7	9,328	12
4	87,8	11,285	8
5	78,5	8,9822	4
6	75,0	12,2573	4
7	92,5	16,415	12
8	70,9	21,958	4
9	96,1	19,356	24
10	93,5	17,433	12
11	91,5	21,642	8
12	78,7	11,488	4

Source. Own elaboration.

Figure 4 ranks vector cases according to the performance of fault location. Figure 5 ranks them based on computational times. Both figures demonstrate that cases involving data from the three phases outperform their respective only faulty-phases cases regarding the two indicators. The only exception is the comparative data of computational effort between cases 10 and 12, where the only fault-phase case (12) obtained a better computational time.

Data including variations in phase magnitudes reported less estimation error of faults in shorter computational time than those consisting of either phasor variations or pre-fault/fault information of the same magnitudes. The previous is attested in comparisons of case 2 against cases 7 and 9. Despite reporting a small difference in

performance, cases 5 and 8 corroborate the superiority of variations in phase magnitudes.

For the two comparative indicators of the study, in-phase data (cases 2 and 5) and case outperform in-line data (cases 3 and 6). However, their combinations (cases 1 and 4) showed an improvement only in fault location.
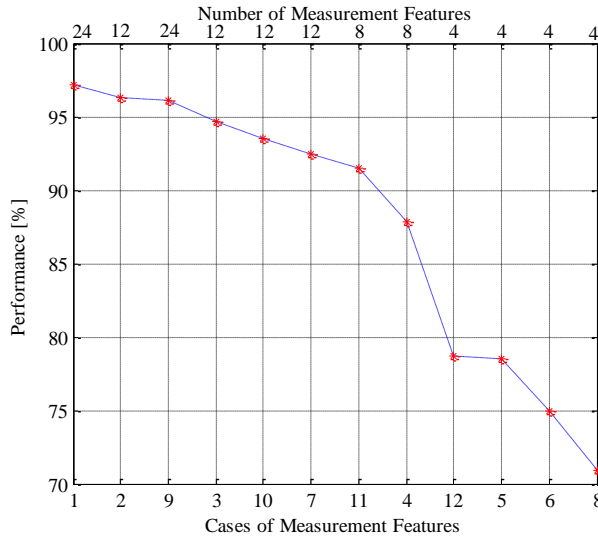

Figure 4. Cases and number of measurement features vs performance values for single phase to ground fault, phase a. Source. Own elaboration.
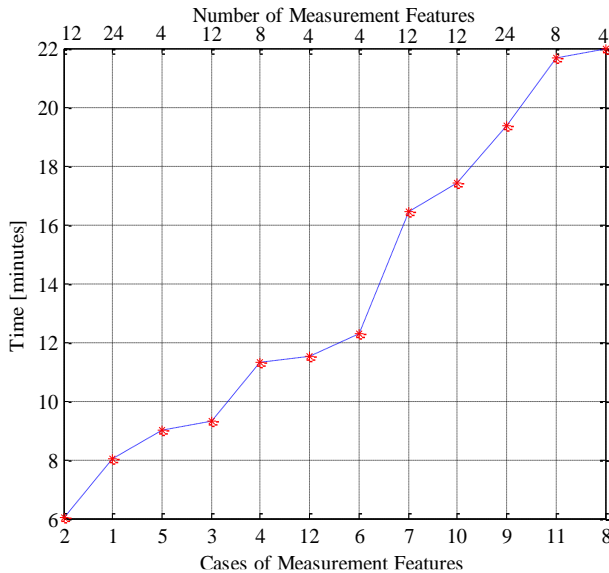

Figure 5. Cases and number of measurement features vs time values for single phase to ground fault, phase a. Source. Own elaboration.

The parameters for each combination of measurement features in Table 3 are determined before carrying out the process of cross-validation. This parameter determination is performed using the automatic method of parameter selection of the reference [10]. The fourth column of Table 5 reveals the computing time to obtain the parameters for some combinations of the measurement features. The cases of Table 5 are compared with Figure 4, and it is observed that, for a better performance, a greater time of parameterization is obtained. However, by analyzing the Figure 5, it is no observed an influential pattern in the time of selecting the parameters and the time of cross-validation.

Table 5. Results summary of some cases of parameterization times.

Cases of Measurement Features	Number of Measurement Features	Measurement Features (attributes: x_i)	Parameter Time [hours]
2	12	dI, dV, $dAngI$, $dAngV$.	5.46
9	24	I_p, I_f, V_p, V_f, $AngI_p$, $AngI_f$, $AngV_p$, $AngV_f$.	4.93
7	12	$|dI|$, $|dV|$, $Ang(dI)$, $Ang(dV)$.	2.12

4. Conclusions

The best performances are achieved in cases 1, 2 and 9. Analyzing the test regarding computational cost, the best cases are 2, 1 and 5. According to these results, it is suggested to use the measurement features of dI, dV, dAngI, dAngV, which corresponds to case 2, since it shows a good performance with the lowest computational cost.

Quality of measure features may increase with the inclusion of two kinds of data: 1) data regarding three phases rather than only the faulty phase and 2) variations in phase magnitude.

The study also found that in-phase data may be more relevant than information related to lines. According to results, computational time is not related to either performance or number of measurement features.

Future works should focus on verifying these findings with another type of attributes of SVM-c (power factor, apparent power or reactance).

Finally, this localization machine helps improve supply continuity indices and thus, enhance the performance of the power system and the customers served.

Comparison of measurement features used as inputs in a learning-based fault location method for power distribution systems

79

References

[1] T. Short, *Electric Power Distribution Handbook*. New York: CRC press, 2003.

[2] A. Bahmanyar, S. Jamali, A. Estebsari, and E. Bompard "A comparison framework for distribution system outage and fault location methods," *Electr. Power Syst. Res.,* vol. 145, pp. 19–34, 2017. doi: 10.1016/j.epsr.2016.12.018

[3] I. D. Serna Suárez, G. Carrillo Caicedo, H. R.Vargas Torres, "Revisión de técnicas de estado estable y transitorio para la localización de fallas en sistemas de distribución," *Rev. UIS Ing.*, Vol. 9, No. 1, pp. 23- 38, 2010.

[4] J. Gutiérrez, J. Mora, S Pérez, "Strategy based on genetic algorithms for an optimal adjust of a support vector machine used for locating faults in power distribution systems," *Rev.Fac.IngUniv. Antioquia*, no. 53, pp. 174-187, 2010.

[5] J. Mora-Florez, V. Barrera-Nunez and G. Carrillo-Caicedo, "Fault Location in Power Distribution Systems Using a Learning Algorithm for Multivariable Data Analysis," in *IEEE Transactions on Power Delivery*, vol. 22, no. 3, pp. 1715-1721, 2007. doi: 10.1109/TPWRD.2006.883021

[6] J. Mora-Flòrez, J. Meléndez, y G. Carrillo-Caicedo, "Comparison of impedance based fault location methods for power distribution systems," *Electr. Power Syst. Res*., vol. 78, no. 4, pp. 657–666, 2008. doi: 10.1016/j.epsr.2007.05.010

[7] A. A. Girgis, C. M. Fallon and D. L. Lubkeman, "A fault location technique for rural distribution feeders," in *IEEE Transactions on Industry Applications*, vol. 29, no. 6, pp. 1170-1175, 1993. doi: 10.1109/28.259729.

[8] S. Shilpa G, H. Mokhlis, H. Illias, A. H. Abu Bakar, y L. Awalin, "Fault Identification in an Unbalanced Distribution System Using Support Vector Machine", *J. Electr. Syst*., vol. 12, no. 4, pp. 786–800, 2016.

[9] V. A. Barrera Núñez, G. Carrillo Caicedo, G. Ordóñez Plata, J. J. Mora Flórez, "Una aplicación de la técnica LAMDA a los índices de continuidad del suministro de energía eléctrica," *Rev. UIS Ing.,* vol. 5, no. 1, pp. 25-36, 2006.

[10] C. Li, C. Lin, B. Kuo and H. Ho, "An Automatic Method for Selecting the Parameter of the Normalized Kernel Function to Support Vector Machines," *2010 International Conference on Technologies and Applications of Artificial Intelligence*, Hsinchu, 2010, pp. 226-232. doi: 10.1109/TAAI.2010.46.

[11] W. Gil, J. Mora, S Pérez, "Analysis of the input data processing for fault location in power distribution systems," *Tecnura*, vol. 18, no. 41, pp. 64 – 75, 2014.

[12] D. Arredondo, J. Mora, L. Román, "Exhaustive Search of Input Characteristics to Improve the Performance of Support Vector Machines for Fault Location," *Energética,* vol. 44, pp. 69-74, 2014.

[13] D.J. Arredondo, W.J. Gil, J.J. Mora, "Methodology for selection of attributes and operating conditions for SVM-Based fault locator's," *Tecnura*, vol. 21, no. 51, pp. 15-26, 2017.

[14] S. Maldonado, R. Weber, "Modelos de Selección de Atributos para Support Vector Machines," *Rev. Ing. de Sist.*, vol. 26, pp. 49-70, 2012.

[15] J. A. Cormane Angarita, H. R. Vargas Torres, G. Ordoñez Plata, "Aplicación de mezcla de distribuciones a la localización de fallas en sistemas de distribución de energía eléctrica," *Rev. UIS Ing.,* vol. 5, No. 1, pp. 49-57, 2006.

[16] G. A. Morales España, H. R. Vargas Torres, J. J. Mora Flórez, "Influencia de la variación en la carga y del tamaño de la zona en la precisión de un localizador de fallas para circuitos de distribución," *Rev. UIS Ing*., vol. 6, no. 1, pp. 47-57, 2007.

[17] M. Brunner, G. Nagy, y O. Wilhelm, "A Tutorial on v-Support Vector Machines," *J. Pers*., vol. 80, no. 4, pp. 796–846, 2012. doi: 10.1111/j.1467-6494.2011.00749.x

[18] J. Mora, S Pérez, "Reducción del tamaño de la zona bajo falla para determinar el desempeño de un localizador de fallas basado en vectores de soporte y aplicado a sistemas de distribución," *Rev. Tecnura*, vol. 10, no. 20, pp.78–89, 2007.

[19] C. Chih-Chung y L. Chih-Jen, "LIBSVM -- A Library for Support Vector Machines". [Online]. Available: https://www.csie.ntu.edu.tw/~cjlin/libsvm/. [Accessed: Jul 25, 2017]

[20] J. J. Mora, J. C. Bedoya and J. Melendez, "Extensive Events Database Development using ATP and Matlab to Fault Location in Power Distribution Systems," *2006 IEEE/PES Transmission & Distribution Conference and Exposition: Latin America,* Caracas, 2006, pp. 1-6. doi: 10.1109/TDCLA.2006.311426

[21] J. Dagenhart, "The 40-ohm ground fault phenomenon," *1999 Rural Electric Power Conference (Cat. No. 99CH36302)*, Indianapolis, IN, USA, 1999, pp. C4/1-C4/3. doi: 10.1109/REPCON.1999.768690