



Tecnura

ISSN: 0123-921X

tecnura.ud@correo.udistrital.edu.co

Universidad Distrital Francisco José de

Caldas

Colombia

Arredondo Arteaga, Debbie Johan; Gil González, Walter Julián; Mora Flórez, Juan José  
Metodología para la selección de atributos y condiciones operativas para la localización  
de fallas basada en la máquina de soporte vectorial

Tecnura, vol. 21, núm. 51, january-march, 2017, pp. 15-26

Universidad Distrital Francisco José de Caldas

Bogotá, Colombia

Disponible en: <http://www.redalyc.org/articulo.oa?id=257050668002>

- ▶ Cómo citar el artículo
- ▶ Número completo
- ▶ Más información del artículo
- ▶ Página de la revista en redalyc.org



Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal  
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

## Metodología para la selección de atributos y condiciones operativas para la localización de fallas basada en la máquina de soporte vectorial

### Methodology for selection of attributes and operating conditions for SVM-Based fault locator's

Debbie Johan Arredondo Arteaga<sup>1</sup>, Walter Julián Gil González<sup>2</sup>, Juan José Mora Flórez<sup>3</sup>

Fecha de recepción: 18 de marzo de 2016

Fecha de aceptación: 23 de noviembre de 2016

**Cómo citar:** Arredondo A., D.J.; Gil-G., W.J. y Mora F., J.J. (2017). Metodología para la selección de atributos y condiciones operativas para la localización de fallas basada en la máquina de soporte vectorial. *Revista Tecnura*, 21(51), 15-26.

#### Resumen

**Contexto:** Las técnicas de localización de fallas se presentan como una alternativa ágil de restauración del servicio eléctrico en las redes de distribución de energía, debido a que los circuitos de distribución son generalmente de gran tamaño y las interrupciones del servicio son comunes. Por tanto, las empresas distribuidoras deben emplear estrategias para cumplir con su servicio oportuno y de alta calidad. Sin embargo, las técnicas de localización son poco robustas y presentan algunas limitaciones en costo computacional y en la descripción matemática de los modelos utilizados.

**Método:** Este artículo está orientado al análisis de las condiciones adecuadas de ajuste y validación de un localizador de fallas para sistemas de distribución, basado en la máquina de soporte vectorial. Con esto es posible determinar el número mínimo de condiciones operativas que permiten alcanzar un buen desempeño con un bajo esfuerzo computacional.

**Resultados:** La metodología propuesta se prueba en un circuito de distribución prototípico rural

de Colombia a 34,5 kV, subdividido en 20 zonas, el cual, ante fallas monofásicas y diferentes condiciones de operativas, permite obtener una base de datos de 630.000 registros. Como resultado, se determina que a partir de 200 condiciones operativas adecuadamente seleccionadas, el localizador mostró un desempeño superior al 98 %.

**Conclusiones:** Es posible mejorar el desempeño de localizadores de fallas basados en la máquina de soporte vectorial (SVM), mediante la selección adecuada de atributos y condiciones operativas, las cuales afectan directamente el rendimiento en términos de desempeño y costo computacional.

**Palabras clave:** falla monofásica, localizador de fallas, máquina de soporte vectorial.

#### Abstract

**Context:** Energy distribution companies must employ strategies to meet their timely and high quality service, and fault-locating techniques represent and agile alternative for restoring the electric service in the power distribution due to the size of distribution

1 Ingeniero electricista, magíster en Ingeniería Eléctrica. Universidad Tecnológica de Pereira (UTP). Pereira, Colombia.  
Contacto: djarredondo@utp.edu.co

2 Ingeniero electricista, magíster en Ingeniería Eléctrica. Universidad Tecnológica de Pereira (UTP). Pereira, Colombia.  
Contacto: wjgil@utp.edu.co

3 Ingeniero electricista, doctor en Ingeniería Eléctrica. Docente titular de la Universidad Tecnológica de Pereira (UTP). Pereira, Colombia.  
Contacto: jjmora@utp.edu.co

services (generally large) and the usual interruptions in the service. However, these techniques are not robust enough and present some limitations in both computational cost and the mathematical description of the models they use.

**Method:** This paper performs an analysis based on a Support Vector Machine for the evaluation of the proper conditions to adjust and validate a fault locator for distribution systems; so that it is possible to determine the minimum number of operating conditions that allow to achieve a good performance with a low computational effort.

**Results:** We tested the proposed methodology in a prototypical distribution circuit, located in a rural area of Colombia. This circuit has a voltage of 34.5

KV and is subdivided in 20 zones. Additionally, the characteristics of the circuit allowed us to obtain a database of 630.000 records of single-phase faults and different operating conditions. As a result, we could determine that the locator showed a performance above 98% with 200 suitable selected operating conditions.

**Conclusions:** It is possible to improve the performance of fault locators based on Support Vector Machine. Specifically, these improvements are achieved by properly selecting optimal operating conditions and attributes, since they directly affect the performance in terms of efficiency and the computational cost.

**Keywords:** fault locator, single phase fault, support vector machine.

## INTRODUCCIÓN

La continuidad del servicio de energía eléctrica ha sido ampliamente estudiada, ya que es considerada un aspecto importante de la calidad en el nuevo entorno de mercado, y debido a su fuerte impacto sobre el servicio al cliente y la repercusión que tiene sobre los ingresos de las empresas relacionadas con energía eléctrica. Una de las causas más relevantes que afectan la continuidad del servicio son las fallas permanentes, las cuales se caracterizan por requerir de la intervención humana para la restauración del servicio. Para lograr que el tiempo de desconexión de los usuarios sea mínimo, se han propuesto metodologías robustas que determinen la posición de la falla, para reducir así el tiempo total de la indisponibilidad.

Dos de las metodologías más usadas en la localización de fallas son los métodos basados en el modelo (MBM) y los métodos basados en el conocimiento (MBC). Los MBM estiman la distancia desde la subestación a la falla, a partir del modelo matemático en la subestación. Aunque estas metodologías presentan gran velocidad en la estimación, tienen errores debidos a la múltiple estimación del sitio de falla, debido a las

características radiales y altamente ramificadas de los sistemas de distribución, tal como se describe en Flórez y Londoño (2007), y Thukaram, Khincha y Vijaynarasimha (2005). Por otro lado los MBC determinan la localización de la falla con la información que proviene de los registros de tensión y corrientes medidos en la subestación. Esta metodología no tiene el problema de la múltiple estimación, aunque sus altos costos computacionales lo hacen poco deseados.

Los MBC se centran en técnicas que tengan la habilidad de reconocer patrones tal como lo hace la máquina de soporte vectorial (SVM). La SVM pertenece a las técnicas de aprendizaje supervisado, su objetivo principal es extraer el conocimiento de las bases de datos; la aplicación de esta técnica al problema de localización de fallas ha tenido gran éxito (Flórez y Londoño, 2007; Mora-Flórez, Meléndez y Carrillo-Caicedo, 2008; Morales-España, Mora-Flórez y Vargas-Torres, 2010). Sin embargo, en las referencias citadas no se consideran la selección adecuada de las condiciones de prueba, como también el número de condiciones operativas para parametrización y entrenamiento, y finalmente no se evalúa el tiempo requerido para la prueba (entrenamiento y validación).

En este artículo se realiza un análisis que permite encontrar las variables que afectan los métodos de localización de fallas basado en máquinas de soporte vectorial, desde la selección de atributos, las constantes de parametrización, el tamaño de la base de datos en función del número de condiciones operativas para el entrenamiento; para así poder determinar un número adecuado en tiempo y condiciones para un sistema de distribución de energía eléctrica.

## FUNDAMENTACIÓN

### Máquina de soporte vectorial (SVM)

Las SVM tienen su origen en la teoría de aprendizaje estadístico la cual, que tiene como característica especial la independencia del conocimiento inicial de la densidad de probabilidad de los datos que se analizan (Moguerza y Muñoz, 2006). Otra ventaja importante de esta técnica de aprendizaje es que solo tiene dos variables de ajuste, la constante de penalización C y el parámetro del kernel. Para un kernel de base radial (RBF) dado por la ecuación (1) (Yao, Hu, Zhang y Jin, 2014).

$$k(\vec{x}, \vec{y}) = e^{-\gamma|\vec{x}-\vec{y}|^2} \quad (1)$$

Donde,  $\vec{x}$  y  $\vec{y}$  corresponden a las entradas de la SVM.

### Búsqueda tabú (TS)

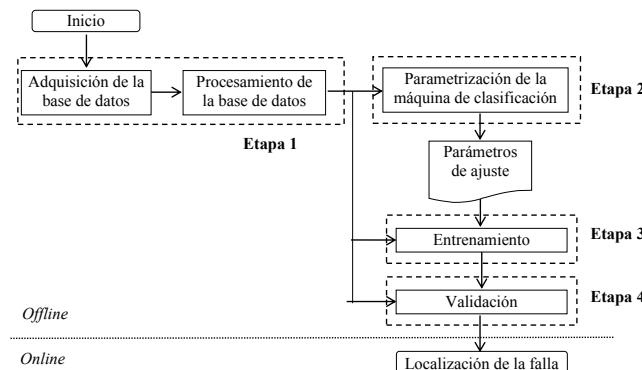
Esta estrategia de optimización metaheurística encuentra soluciones factibles a problemas complejos con un coste computacional aceptable en los equipos de cómputo modernos.

Esta técnica considera los registros de búsqueda con una estrategia de memoria, que le permite explorar adecuadamente el espacio de solución (Yao et al., 2014). La lista tabú sirve para modelar la memoria y facilita el seguimiento y exclusión de las soluciones ya evaluadas. Cada paso o ejecución del algoritmo, permite encontrar una solución y la mejor de ellas se incluye en la lista.

## METODOLOGÍA

La metodología general utilizada para encontrar las variables que tienen mayor influencia en el desempeño del localizador, se muestran en la figura 1.

En la etapa 1 se adquieren los registros de falla del sistema de potencia y se procesan los datos necesarios para que los MBC tengan un mejor rendimiento. En la etapa 2 se seleccionan los parámetros de la SVM usando la técnica metaheurística de búsqueda tabú. En la etapa 3 se entrena la SVM para que relacione la localización de las fallas con los registros de entrada, mediante la creación de los vectores de soporte. Finalmente, en la etapa 4 se realiza la validación del desempeño del localizador.



**Figura 1.** Esquema general del método de clasificación

**Fuente:** elaboración propia.

## Etapa 1

### *Adquisición de datos*

Los registros de falla utilizados se obtienen a partir de la simulación intensiva del sistema de distribución, utilizando una herramienta desarrollada en ATP y MATLAB. Las fallas utilizadas consideran diferentes situaciones operativas, valores diferentes de resistencia de falla y varios tipo de fallas (Hernández, Londoño, & Flórez, 2009).

### *Zonificación de la red de distribución*

Como la SVM es una técnica de aprendizaje supervisado, se requiere que en el entrenamiento se relacionen tanto los datos de falla como la clase a la cual pertenecen. La clase en este caso se define como una de las subdivisiones del sistema de distribución bajo análisis. Una subdivisión o zona se define como un subsistema radial, para reducir el problema de múltiple estimación del sitio en falla (Zapata-Tapasco, Mora-Flórez y Pérez-Londoño, 2014). El tamaño de esta subdivisión puede variar según el criterio del operador del sistema de potencia y las necesidades específicas de cada sistema. La división en zonas que se propone en este artículo se realizó mediante una herramienta automática de zonificación, propuesta por Zapata (2013).

### *Procesamiento de los datos de entrada*

El procesamiento de datos es una etapa de suma importancia en las técnicas de aprendizaje, debido a que los algoritmos de optimización de fondo requieren un tiempo menor de convergencia si los datos son procesados; en la fase de procesamiento se realizan actividades como la extracción de características, normalización, filtrado y selección de los datos (Kotsiantis, Kanellopoulos y Pintelas, 2006).

La selección de los atributos es una de las fases de la etapa de procesamiento, donde se toman las características de las fallas las cuales contienen la mayor cantidad de información posible. Estos atributos son extraídos de los fasores de tensión y corriente medidos en una subestación aguas arriba, esta subestación alimenta el ramal de distribución

bajo falla. En este artículo, los atributos utilizados corresponden a la selección de las mejores combinaciones propuestas por Arredondo, Mora y Román (2014). Además, es necesario relacionar cada falla de acuerdo con la ubicación en la cual ocurrió, a este proceso se le conoce como *asignación de etiquetas* (Zapata-Tapasco et al., 2014).

Ya que es posible encontrar un atributo (o entre ellos) con grandes diferencias de magnitudes asociadas a las fallas (Kotsiantis, Kanellopoulos y Pintelas, 2006), es necesario el proceso de estandarización de la base de datos, en esta investigación fue implementada la técnica de normalización *Min-max*, que ha presentado buen desempeño en el problema de localización de fallas (Zapata-Tapasco, Mora-Flórez y Pérez-Londoño, 2014; Gil, Mora y Pérez, 2014; Morales-España, Mora-Flórez y Vargas-Torres, 2010).

Esta técnica es una de las más simples, los valores mínimos y máximos de la base de datos se desplazan a los valores 0 y 1, respectivamente, y todos los datos se transforman en el rango {0,1} (Safadi, Derbas y Quénnot, 2015), utilizando la ecuación (2).

$$s'_{ij} = \frac{s_{ij} - \min(s_j)}{\max(s_j) - \min(s_j)} \quad (2)$$

Donde,  $s'_{ij}$  es el dato  $i$  transformado del conjunto de datos  $j$ ;  $s_{ij}$  es el dato  $i$  original del conjunto de datos  $j$ ;  $\min(s_j)$  es el valor mínimo del conjunto de datos  $j$ ;  $\max(s_j)$  es el valor máximo del conjunto de datos  $j$ .

## Etapa 2

En esta se realiza la selección de los parámetros de penalización y dimensionalidad de la técnica de aprendizaje ( $C$  y  $\gamma$ ); dicha selección no es un proceso trivial, debido al espacio de selección y los altísimos costos computacionales. Debido a esto, es necesario implementar una técnica metaheurística basada en la exploración del espacio de soluciones, logrando alcanzar puntos de alta calidad.

En este artículo se implementó la técnica llamada *búsqueda tabú*, donde el error de validación cruzada se utiliza como la función objetivo, y así el algoritmo evoluciona hasta encontrar el menor error de validación.

La validación cruzada es el proceso en el cual se exploran los cúmulos de información de la base

de datos, el cual consiste en dividir esta última en partes iguales, para así entrenar la técnica con  $n-1$  subconjuntos y validar con el subconjunto restante, así hasta alcanzar los  $n$  entrenamientos y validaciones, al final se promedian los errores encontrados y se obtiene un solo error de validación (Zapata-Tapasco, Mora-Flórez y Pérez-Londoño, 2014).

$$\text{Error de validación} = \frac{\text{Número de datos mal clasificados}}{\text{Número de datos de prueba}} \cdot 100\% \quad (3)$$

### Etapa 3

En esta etapa se utilizan diferentes condiciones operativas del sistema de potencia, las cuales consideran variación en los parámetros del circuito. Se hicieron pruebas considerando diferentes conjuntos de entrenamiento (variando el número de condiciones operativas), tomando en cuenta tiempo y precisión. El entrenamiento puede determinar el desempeño de la máquina de soporte tanto en tiempo como en error, debido a esto es importante encontrar un número mínimo de condiciones operativas que sean suficientes en el entrenamiento de la SVM y obtener un buen desempeño del clasificador, y así la estimación de zona en falla se reduzca considerablemente en el tiempo de ejecución. Adicionalmente, al trabajar con menos

condiciones operativas se evitan sobreentrenamientos, los cuales no mejoran el desempeño del localizador, pero sí aumentan el tiempo de cálculo de esta metodología (uno de los factores más importantes en el trabajo con MBC es el tiempo de cálculo, tiempos elevados la hacen poco deseable en comparación con metodologías como las MBM).

### Etapa 4

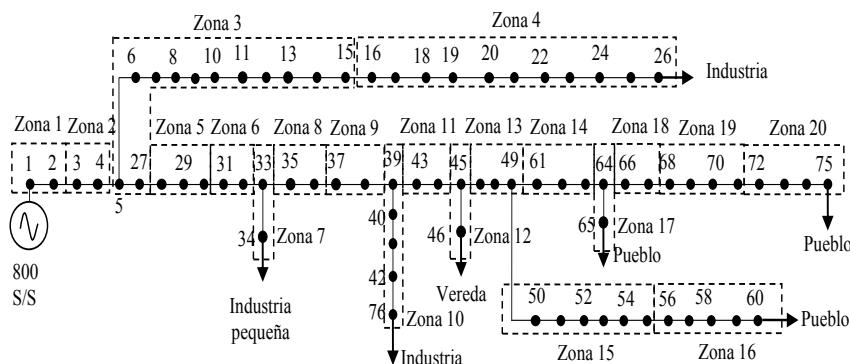
Finalmente, esta etapa consiste en evaluar el rendimiento del localizador, que se realiza a partir del cálculo de la precisión del método ante datos no considerados en las etapas de parametrización y entrenamiento. La precisión se calcula como se muestra en la ecuación (4).

$$\text{Precisión} = \frac{\text{Número de datos bien clasificados}}{\text{Número de datos de prueba}} \cdot 100\% \quad (4)$$

## RESULTADOS

La metodología propuesta se prueba en un circuito prototípico rural presentado en Osorio (2014), que opera a 34,5 kV. Este circuito fue diseñado con el fin de considerar las características generales de las redes de distribución en Colombia, como se presentan en las zonas rurales de los departamentos de Boyacá y Cundinamarca. Además, presenta cargas industriales y veredas donde su comportamiento es predominantemente inductivo, laterales

trifásicos y múltiples calibres de conductor, las estructuras, normas técnicas y su descripción detallada se describe en Osorio (2014). La zonificación del circuito se obtiene mediante un algoritmo de zonificación automática, el cual está basado en la técnica metaheurística *tabú* y una zonificación inicial hecha por ingenieros que conocen el circuito en detalle. El algoritmo se encuentra descrito en Zapata (2013) (el circuito se divide en 20 zonas), tal como se muestra en la figura 2.

**Figura 2.** Red 34,5kV rural

**Fuente:** elaboración propia.

Los escenarios considerados para probar la metodología solo contienen fallas monofásicas, debido a la alta carga computacional de las pruebas. Además, este tipo de fallas representan aproximadamente un 70 % de las fallas que se presentan en un circuito de distribución (Mora, 2007). Se simulan diferentes resistencias de fallas que varían entre  $0,05 \Omega$  hasta  $40 \Omega$  en paso de  $4 \Omega$ , los cuales son valores comunes en sistemas de distribución (Dagenhart, 2000). Adicionalmente, se simularon 1400 condiciones operativas del sistema de potencia, para tener una representación real del circuito. Para cada condición operativa se simulan fallas en todos los nodos, de donde se obtienen 450 fallas por condición operativa, que conforman una base de datos de 630.000 fallas. Las condiciones operativas se obtienen con la ayuda de la herramienta desarrollada por Alzate-González, Mora-Flórez y Pérez-Londoño (2014). Se utiliza la técnica de muestreo *Latin Hypercube*, la cual explora todo el espacio de soluciones, con lo cual se considera un gran número de condiciones de operación del circuito. La herramienta permite hacer variaciones de los parámetros del circuito como: magnitud y ángulo de la tensión de la fuente, carga promedio, factor de potencia, longitud de la línea y resistividad del terreno. Los rangos de variación considerados para cada parámetro se presentan en la tabla 1, estas variaciones se realizan de acuerdo con los valores nominales del circuito, por ejemplo: si se

tiene un factor de potencia nominal de una carga de 0,9 en atraso, la herramienta puede variar este valor entre 0,88 y 0,92.

**Tabla 1.** Rangos de la variación de los parámetros del circuito

Parámetros	Rango de incertidumbre	
	Mínimo	Máximo
Magnitud de tensión en la fuente	0,95 p.u	1,10 p.u
Ángulo de tensión en la fuente	-3,4°	3,4°
Carga promedio	10 %	150 %
Factor de potencia	-0,02	0,02
Longitud del conductor	95 %	105 %
Resistividad del terreno	80 Ωm	120 Ωm

**Fuente:** elaboración propia.

Las variaciones como la magnitud de la tensión en la fuente, el ángulo de la tensión en la fuente, son globales, esto quiere decir que solo se realizan una vez por circuito. El factor de potencia y la longitud de la línea también fueron variaciones globales, es decir, para cada factor de potencia en la carga y cada segmento de línea se realizó el mismo porcentaje de variación. Sin embargo, la magnitud de cada carga del sistema varió de forma individual cada una, con la idea de representar condiciones más severas de las que se presentan en un circuito real.

La combinación de atributos utilizados fue determinada con una búsqueda exhaustiva, como la presentada en Arredondo, Mora y Román (2014). Para este circuito las mejores tres combinaciones se presentan en la tabla 2.

**Tabla 2.** Mejores combinaciones de atributos

Número de combinación	Atributos
Combinación 1	$dV$ , $dl$ , $d\theta_V$
Combinación 2	$dV$ , $dl$ , $d\theta_I$
Combinación 3	$dV_I$ , $dl_I$ , $d\theta_{VI}$ , $d\theta_{II}$

**Fuente:** elaboración propia.

### Pruebas a condición nominal en la parametrización

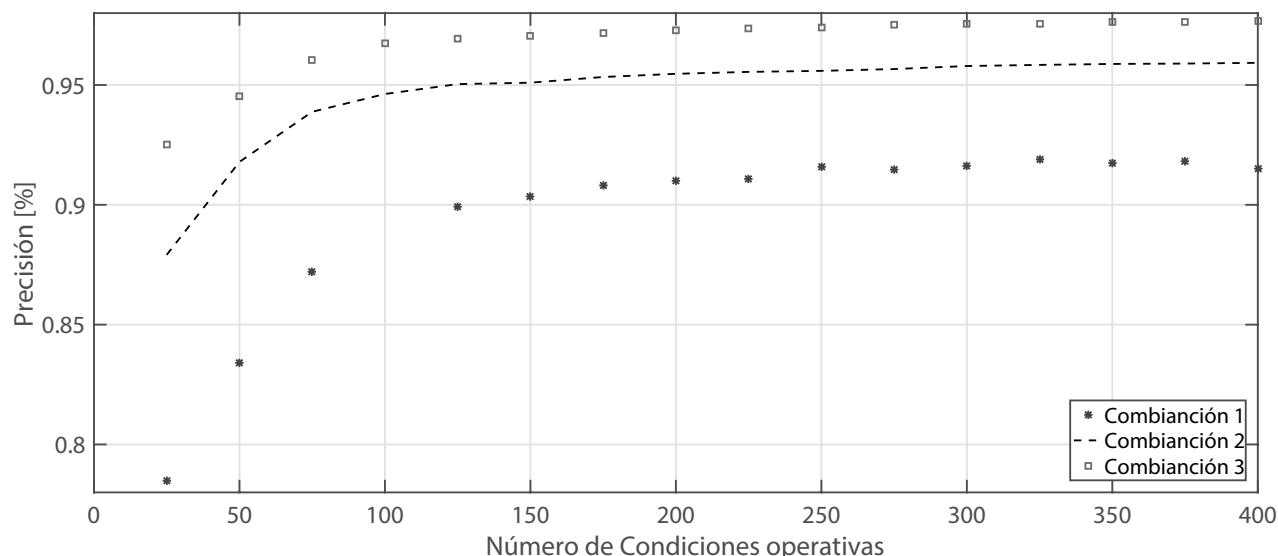
Esta prueba se parametrizó con fallas simuladas a condición nominal del circuito, con un total de 450

fallas. En el entrenamiento y la validación se utilizaron todas las condiciones operativas simuladas.

La etapa de entrenamiento se realizó con diferentes conjuntos de condiciones operativas, utilizando inicialmente 25 e incrementando en cada prueba 25 condiciones operativas hasta 400 (esto se realiza para cada combinación de descriptores utilizada). Además, para determinar el número máximo de condiciones operativas, se tuvo en cuenta que el desempeño del clasificador no aumentará considerablemente y que el costo computacional fuera muy alto.

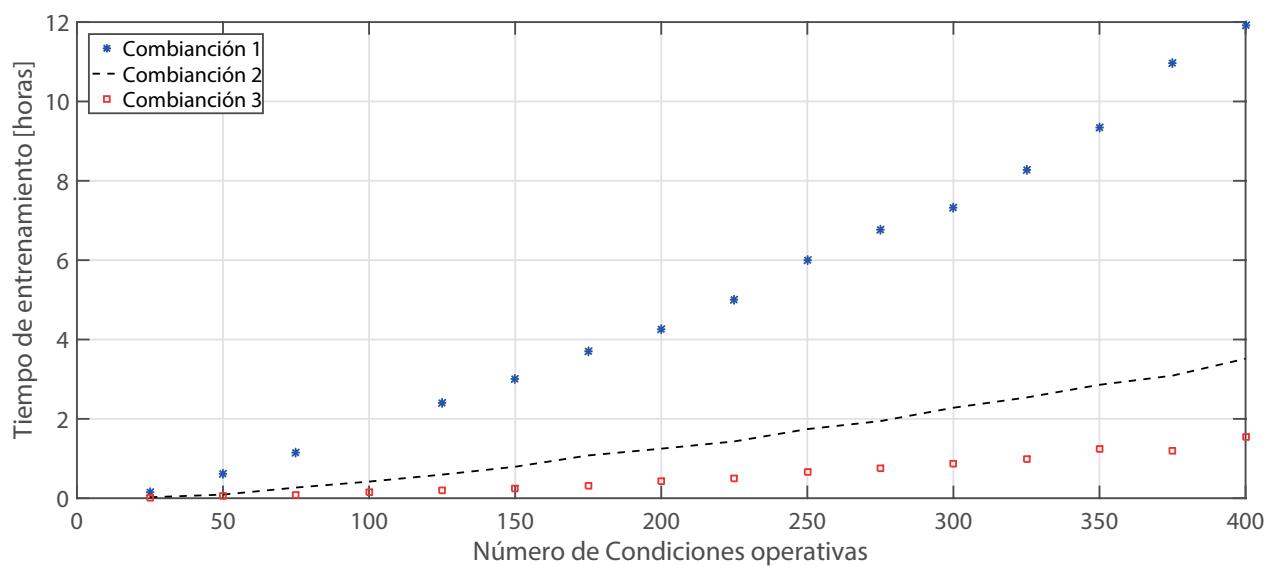
La etapa de validación se realizó con 1000 condiciones operativas para un total de 450.000 fallas (sin incluir las condiciones operativas ya utilizadas en la parametrización y entrenamiento).

En las figuras 3 y 4 se muestra el desempeño para cada combinación de atributos y el tiempo de entrenamiento para cada número condiciones de operativas que se trabajó, respectivamente.



**Figura 3.** Número de condiciones operativas de entrenamiento vs. desempeño para cada combinación de atributos

**Fuente:** elaboración propia.



**Figura 4.** Número de condiciones operativas de entrenamiento vs. tiempo de entrenamiento para cada combinación de atributos

**Fuente:** elaboración propia.

Se puede observar en las figuras 3 y 4, que la mejor combinación de atributos es la denominada combinación 3, la cual alcanza un mejor desempeño en la validación como en el tiempo de entrenamiento. Además, se puede observar en la figura 3 que alrededor de 200 condiciones operativas el desempeño del localizador no aumenta significativamente, en cambio el tiempo en el entrenamiento crece, lo cual es poco deseable.

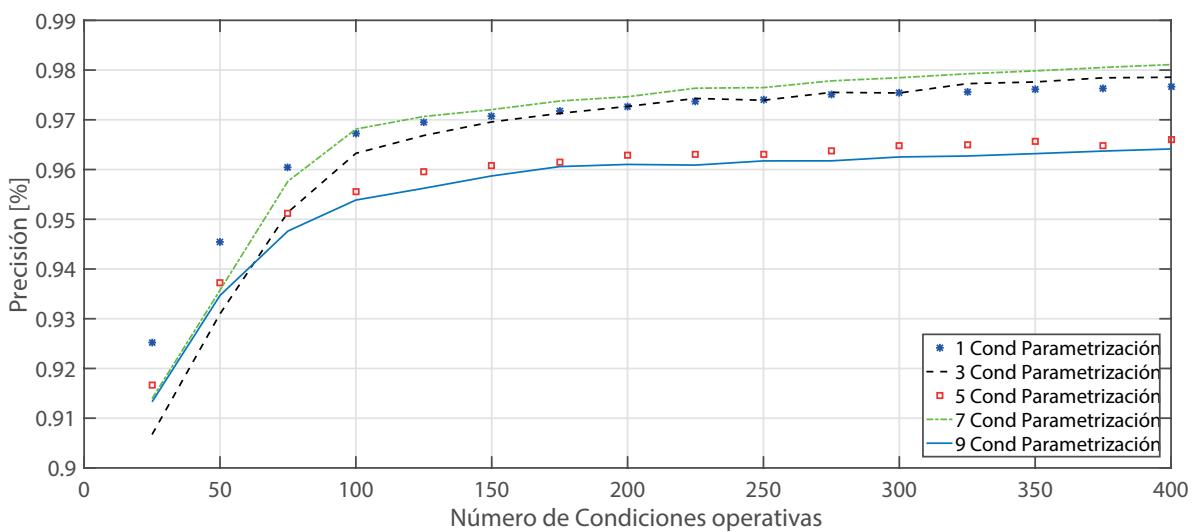
### Pruebas a diferentes condiciones de operación en la parametrización

En esta prueba se busca determinar el número de condiciones operativas en la parametrización, para lograr un mayor desempeño del localizador tanto en rendimiento como en el tiempo de entrenamiento, la prueba solo se realizó para la combinación 3, debido a que fue la que presentó mejores tiempos y desempeños diferentes en el localizador resultados en la prueba anterior. Esta prueba se realizó aumentando de dos en dos las condiciones

operativas, incluyendo siempre la condición nominal del circuito, hasta que no se observara mejora en el desempeño del localizador. Se notó que a partir de nueve condiciones operativas en la parametrización el localizador no presentó mejoría. El entrenamiento y la validación se realizaron igual que en la prueba anterior.

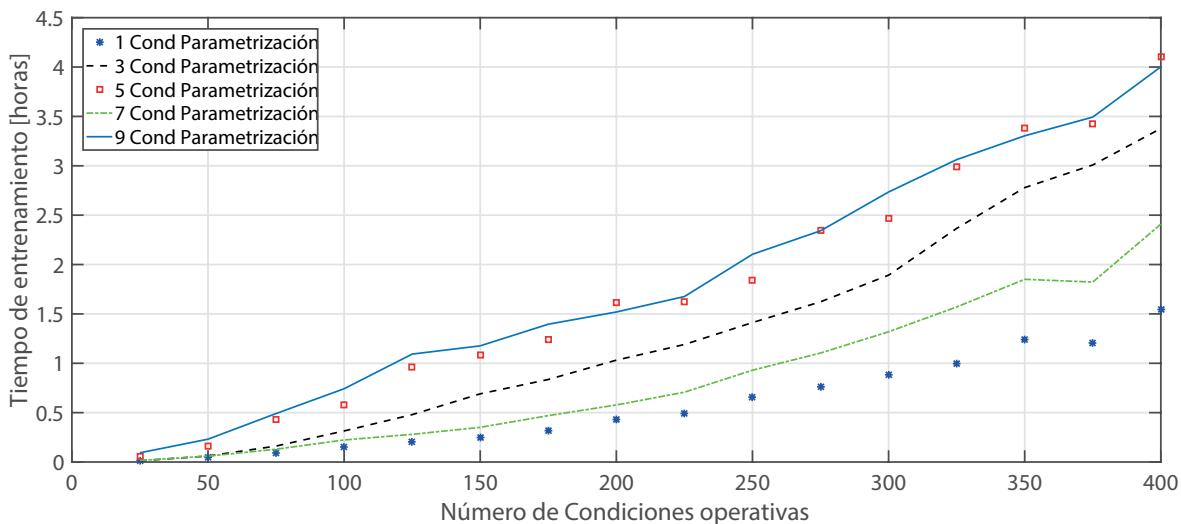
En la figura 5 se muestran los desempeños para cada número de condiciones operativas en la parametrización (No. cond. parametrización) y en la figura 6 se muestran los tiempos de entrenamiento para cada caso.

Se puede notar en las figuras 5 y 6 que la prueba realizada para las diferentes condiciones de parametrización (1, 3, 5, 7, 9) generan tiempos y desempeños diferentes en el localizador, al usar muchas más condiciones en la parametrización se puede notar que es poco eficiente dado el alto costo computacional y el bajo rendimiento de la máquina. Los mejores desempeños se encuentran con siete condiciones operativas, aunque en tiempo de entrenamiento sea un poco menos



**Figura 5.** Número de condiciones operativas de entrenamiento vs. desempeño para cada parametrización

Fuente: elaboración propia.



**Figura 6.** Número de condiciones operativas de entrenamiento vs. tiempo de entrenamiento para cada parametrización

Fuente: elaboración propia.

eficaz que la parametrización con una condición o condición nominal; seleccionar cualquiera de las dos condiciones es una buena elección, aunque es preferible seleccionar la condición nominal dado que contiene menos datos. Se nota que esta prueba no se presentó una tendencia, debido que las pruebas para 1, 3 y 7 condiciones operativas

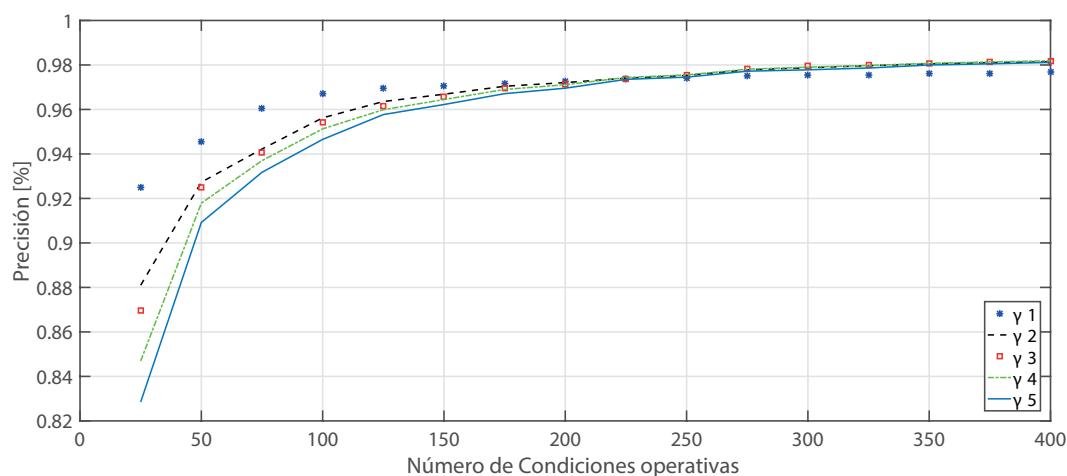
presentan mejores resultados que en el caso 5 y 9. Estos resultados pueden explicarse debido a que las condiciones que presentan buenos resultados representan mucho mejor el comportamiento del circuito, por tanto, se esperaba que la prueba a condición nominal representaría el circuito y sus zonas de falla adecuadamente.

Finalmente, se trata de disminuir el rango de gamma, ya que este parámetro afecta en gran medida el desempeño de la SVM y, por ende, el del localizador. En esta parte se trató de discretizar el rango de gamma en rangos pequeños como se muestra en la tabla 3, y así poder disminuir el tiempo de parametrización y obtener un mejor desempeño con menores esfuerzos computacionales. Esta prueba se realizó solo para la combinación 3 y utilizando siete condiciones operativas en el entrenamiento.

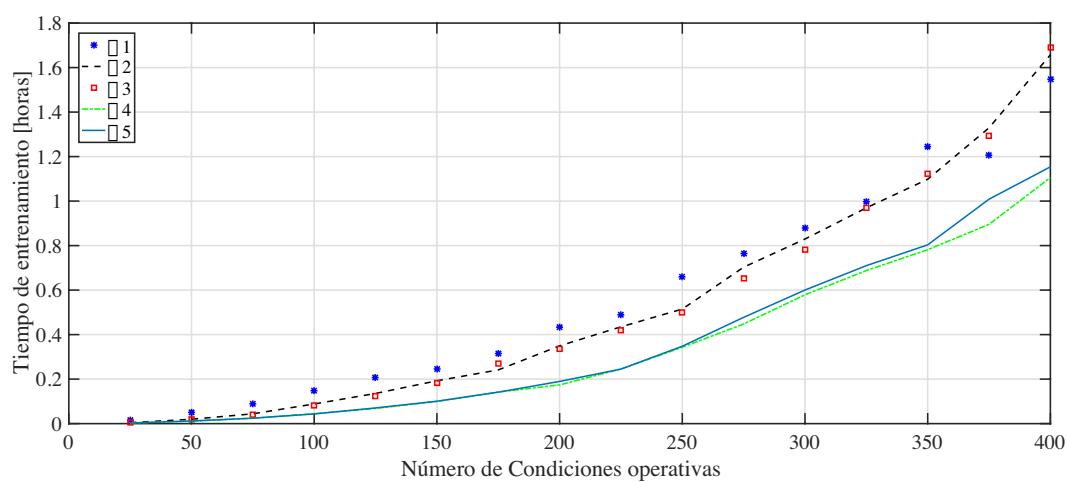
**Tabla 3.** Rangos de variación de gamma

$\gamma$	Rango
1	0-8
2	8-16
3	16-24
4	24-32
5	32-40

**Fuente:** elaboración propia.

**Figura 7.** Desempeño de la máquina con respecto a la variación de gamma

**Fuente:** elaboración propia.

**Figura 8.** Tiempo de entrenamiento con respecto a la variación de gamma.

**Fuente:** elaboración propia.

En la figura 7 se muestran los desempeños para cada rango de gamma, y en la figura 8, los tiempos de entrenamiento para cada caso.

Como se puede observar en la figura 7, la variación del parámetro gamma no afecta significativamente el desempeño, en todos los casos el comportamiento de localizador es similar. Sin embargo, los tiempos de entrenamiento sí varían para cada rango de gamma como se puede notar en la figura 8, siendo para los rangos de gamma 4 y 5 los mejores resultados en tiempo de simulación. Esta variación de los tiempos se debe a que, para algunos rangos de  $\gamma$ , la SVM pueda separar mejor los datos y esto se ve reflejado en el tiempo de entrenamiento.

## CONCLUSIONES

La selección adecuada de los datos de entrenamiento es un paso muy importante en la SVM, debido a que afecta directamente el rendimiento en cuanto a desempeño y costo computacional. En este artículo se demuestra la influencia del conjunto de datos de entrenamiento para un localizador de fallas basado en SVM, considerando tres combinaciones de atributos, variaciones del rango del parámetro del kernel durante la parametrización y el tamaño de la base de datos para entrenamiento. En general, el localizador presentó un buen desempeño para todas las combinaciones utilizadas; sin embargo, la combinación tres presentó el mejor desempeño con precisiones hasta 98,21 % para todas las pruebas propuestas. Adicionalmente, cuando se considera la combinación tres, se requiere un menor esfuerzo computacional en todas las pruebas.

También, se observa que el número de condiciones operativas en el entrenamiento es el factor que más influye en el desempeño del localizador. Se aprecia que a partir de 200 condiciones operativas el desempeño no varía significativamente, sin embargo, el tiempo computacional crece rápidamente.

Finalmente, se puede notar que el localizador tiene un alto grado de confiabilidad, debido a que las variaciones realizadas al circuito de prueba son exhaustivas con la idea de simular cualquier condición operativa que se pueda presentar en un circuito real. Además, en todos los casos trabajados la base de datos de prueba siempre es mayor que la base de prueba en los entrenamientos (un 60 % más grande en el peor de los casos). Estos buenos desempeños también suceden debido a que la SVM tiene una alta capacidad de generalización, lo que implica que provee buenos resultados ante datos desconocidos, el cual la hace útil en sistemas de distribución reales.

## REFERENCIAS BIBLIOGRÁFICAS

- Alzate-González, N.; Mora-Flórez, J. y Pérez-Londoño, S. (2014). Methodology and software for sensitivity analysis of fault locators. In 2014 IEEE PES Transmission & Distribution Conference and Exposition–Latin America (PES T&D-LA) (pp. 1–6). Medellín: IEEE. <http://doi.org/10.1109/TDC-LA.2014.6955261>
- Arredondo, D.; Mora, J. y Román, L. (2014). Búsqueda exhaustiva de descriptores para mejorar el desempeño de las máquinas de soporte vectorial en localización de fallas. *Energética*, 44, 69–74. Recuperado de: <http://www.bdigital.unal.edu.co/46935/1/45314-237156-1-PB.pdf>
- Dagenhart, J. (2000). The 40- $\Omega$  ground-fault phenomenon. *IEEE Transactions on Industry Applications*, 36(1), 30–32. <http://doi.org/10.1109/28.821792>
- Flórez, J.J.M. y Londoño, S.M.P. (2007). 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. *Revista Tecnura*, 10(20), 78–89.
- Gil, W.J.; Mora, J.J. y Pérez, S.M. (2014). Análisis del procesamiento de los datos de entrada para un localizador de fallas en sistemas de distribución. *Revista Tecnura*, 18(41), 64–75. <http://doi.org/http://dx.doi.org/10.14483/udistrital.jour.tecnura.2014.3.a05>
- Hernández, L.P.; Londoño, S.P. y Flórez, J.M. (2009). Diseño de una herramienta eficiente de simulación

- automática de fallas en sistemas eléctricos de potencia. *Revista DYNA*, 164(164), 178-188. <http://doi.org/0012-7353>
- Kotsiantis, S.B.; Kanellopoulos, D. y Pintelas, P. (2006). Data Preprocessing for Supervised Learning. *International Journal of Computer Science*, 1(2), 111–117. Recuperado de: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.132.5127&rep=rep1&type=pdf>
- Mora-Flórez, J.; Meléndez, J. y Carrillo-Caicedo, G. (2008). Comparison of impedance based fault location methods for power distribution systems. *Electric Power Systems Research*, 78(4), 657–666. <http://doi.org/10.1016/j.epsr.2007.05.010>
- Mora Flórez, J.J. (2007). *Localización de fallas en sistemas de distribución de energía eléctrica usando métodos basados en el modelo y métodos basados en el conocimiento*. Tesis doctoral. Girona: Universitat de Girona.
- Morales-España, G.; Mora-Flores, J. y Vargas-Torres, H. (2010). Fault location method based on the determination of the minimum fault reactance for uncertainty loaded and unbalanced power distribution systems. In 2010 IEEE/PES Transmission and Distribution Conference and Exposition: Latin America (T&D-LA) (pp. 803–809). IEEE. <http://doi.org/10.1109/TDC-LA.2010.5762977>
- Moguerza, J.M. y Muñoz, A. (2006). Support Vector Machines with Applications. *Statistical Science*, 21(3), 322–336. <http://doi.org/10.1214/088342306000000493>
- Osorio, J. (2014). *Diseño e implementación en ATP de redes de distribución prototipo de media tensión en Colombia para pruebas de localización de fallas*. Pereira: Universidad Tecnológica de Pereira. Recuperado de: <http://repositorio.utp.edu.co/dspace/bitstream/handle/11059/5191/621319O83.pdf?sequence=1>
- Safadi, B.; Derbas, N. y Quénnot, G. (2015). Descriptor optimization for multimedia indexing and retrieval. *Multimedia Tools and Applications*, 74(4), 1267–1290. <http://doi.org/10.1007/s11042-014-2071-6>
- Thukaram, D.; Khincha, H.P. y Vijaynarasimha, H.P. (2005). Artificial neural network and support vector machine approach for locating faults in radial distribution systems. *IEEE Transactions on Power Delivery*, 20(2), 710–721. <http://doi.org/10.1109/TPWRD.2005.844307>
- Yao, B.; Hu, P.; Zhang, M. y Jin, M. (2014). A support vector machine with the tabu search algorithm for freeway incident detection. *Int. J. Appl. Math. Comput. Sci.*, 24(2), 397–404. <http://doi.org/10.2478/amcs-2014-0030>
- Zapata, A. (2013). *Implementación y comparación de técnicas de localización de fallas en sistemas de distribución basadas en minería de datos*. Pererira: Universidad Tecnológica de Pereira. Recuperado de: <http://repositorio.utp.edu.co/dspace/bitstream/handle/11059/3821/621319Z35.pdf?sequence=1>
- Zapata-Tapasco, A.; Mora-Flórez, J. y Pérez-Londoño, S. (2014). Metodología híbrida basada en el regresor knn y el clasificador boosting para localizar fallas en sistemas de distribución. *Ingeniería y Competitividad*, 177(2), 165–177.

