

Búsqueda exhaustiva de descriptores para mejorar el desempeño de las máquinas de soporte vectorial en localización de fallas

Exhaustive Search of Input Characteristics to Improve the Performance of Support Vector Machines for Fault Location

Debbie Johan Arredondo^{a*}, Juan José Mora^a Laura Melissa Román^a

RESUMEN

En este documento se presenta una metodología para encontrar los mejores datos de entrada o descriptores extraídos de las medidas de tensión y corriente de un sistema eléctrico de potencia, los cuales serán utilizados en una máquina de soporte vectorial para localización óptima de fallas. Este problema es importante debido que el tiempo de reposición de los sistemas en falla, deteriora considerablemente la calidad del servicio e incrementa el costo de la operación del sistema de potencia.

La técnica que se propone en este documento, para resolver el problema de localización de fallas, es una búsqueda exhaustiva de los descriptores que presentan altos índices de desempeño de la máquina de soporte vectorial.

En este documento se presentan resultados que muestran cómo el desempeño de la máquina mejora considerando como variables de evaluación al tiempo y al error de validación.

Finalmente, las pruebas se realizan en el circuito de prueba IEEE 34 nodos. Los parámetros, la zonificación recomendada, técnica de normalización se obtienen de estudios previos.

PALABRAS CLAVE

Descriptores; máquina de soporte vectorial; fallas; localización de fallas; sistemas eléctricos de potencia; búsqueda exhaustiva.

ABSTRACT

This document presents a methodology to find the best input data (descriptors), obtained from the measurement of voltage and current of a power electrical system, which are used jointly with support vector machines for an optimal fault location. This problem is important because the reposition time of the faulted systems, highly deteriorate the service quality and increment the cost of the power system operation. The proposed technique is an exhaustive search of the inputs (descriptors), which give high performance indexes in the support vector machine. The results that evidence the performance in time and validation error of the machine are presented in this paper.

Finally, the different tests are performed in a IEEE 34 nodes power system. The used parameters, the zone division and the normalizing strategy were obtained from previous studies.

KEYWORDS

Descriptors; support vector machines; fault; fault location; electrical power network and exhaustive search.

Energética 44, diciembre (2014), pp.69-74

ISSN 0120-9833 (impreso)
ISSN 2357 - 612X (en línea)
www.revistas.unal.edu.co/energetica
© Derechos Patrimoniales
Universidad Nacional de Colombia



1. INTRODUCCIÓN

La calidad del servicio eléctrico se ve afectada de múltiples formas, debido a los tipos de fallas y a los diferentes dispositivos que componen el sistema eléctrico. Esto es un gran problema para los operadores del sistema, dado que una contingencia conlleva muchas pérdidas de dinero tanto para los usuarios como al operador.

El problema más crítico son las fallas permanentes, las cuales hacen que se presente una interrupción del servicio. Normalmente, el operador de red debe encontrar la falla para reconfigurar el sistema y reparar la falla, para así restablecer la continuidad del servicio lo más rápido posible para evitar mayores costos. En sistemas de distribución este problema es complejo debido a las características de la red; dada esta problemática la localización de fallas es un tema ampliamente estudiado, dado a que ayuda a reducir los costos globales de la distribución de energía hasta los usuarios finales [Lawrence, 1992].

Las técnicas para localizar fallas se dividen en dos grandes grupos, el primero se denomina métodos basados en el modelo (MBM) y se enfatizan en el modelamiento matemático del circuito bajo falla, utilizando las mediciones de tensión y corriente en la subestación como se muestra en [Mora, 2008; Morales, 2009]. Estos métodos presentan buenos resultados en la estimación de la distancia aproximada pero debido a la topología no radial de los sistemas de transmisión y distribución se tienen problemas de múltiple estimación [Morales, 2009]. Por otra parte, los métodos basados en el conocimiento MBC se basan en el aprendizaje de máquina de la respuesta del circuito bajo falla, evidenciado en las medidas de tensión y corriente. Estas técnicas permiten eliminar el problema de múltiple estimación presentados por los métodos basados en el modelo [Gutiérrez, et al., 2010; Thukaram, 2005].

Las máquinas de soporte vectorial aplicadas como localizadores de fallas en el sistema eléctrico, permiten encontrar resultados de buena calidad (localizar la zona del sistema donde ha ocurrido un problema - falla). Dado a que las SVM son unas herramientas que hacen parte de un área de estudio de la minería de datos y aprendizaje de máquina, el desarrollo de estas metodologías es de gran utilidad para el mejoramiento de los localizadores de fallas.

Las SVM son técnicas que permiten extraer el conocimiento de las bases de datos para utilizarlo de manera adecuada. Sin embargo, es computacionalmente costoso debido a que requiere de un gran volumen de datos para representar adecuadamente el problema y lograr resultados satisfactorios de desempeño. Trabajos como [Maldonado, 2007; Maldonado, 2012] muestran

como alcanzar mejores resultados con nuevas metodologías, las cuales involucran técnicas metaheurísticas para determinar errores de estimación menores y normalizaciones de los datos de entrada.

En este documento se presenta una estrategia que permite encontrar la mejor combinación de descriptores, los cuales sean fácilmente separables por la metodología de aprendizaje, para generar un alto desempeño en cuanto al problema de localización óptima de fallas.

2 METODOLOGÍA

El aspecto teórico de las máquinas de soporte vectorial se fundamenta en la teoría de aprendizaje estadístico [Moguerza, 2006] y presenta grandes ventajas debido a que no necesita gran entendimiento de la densidad de probabilidad de los datos y con solo dos datos de penalización se puede llegar a un entrenamiento adecuado.

La Ecuación (1) se conoce como la función kernel y es fundamental en la máquina de soporte vectorial dado a que esta función es la que logra separar los datos de forma que la máquina pueda identificar adecuadamente los datos de validación.

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

Donde:

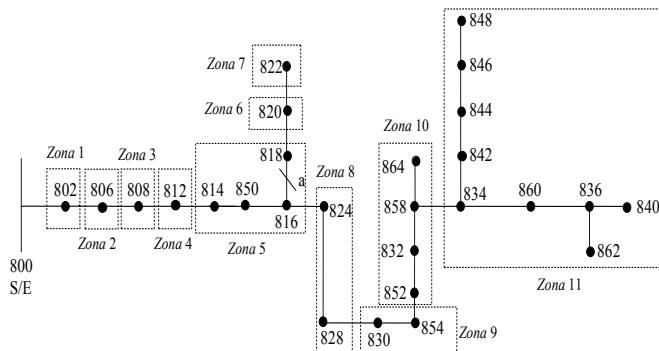
\bar{x} y \bar{y} son los datos de entrada y \cdot es uno de los factores de penalización.

Los datos de un circuito que pueden ser ingresados en la máquina de soporte vectorial para hallar una zona de falla pueden ser: potencia aparente, tensión de fase o línea, corrientes de fase o línea, impedancias, ángulos de corriente o tensión, frecuencia, entre otros. Escoger descriptores adecuados puede aumentar la eficiencia de la máquina tal como se demuestra en [Maldonado, 2012]. El desempeño de la máquina de soporte vectorial depende de diferentes parámetros, como ya se ha dicho los datos de penalización y los descriptores hacen una parte fundamental del problema, pero no son las únicas variables que afectan el resultado final. Por lo anterior, algunos autores [Mora, 2006; Gil, 2011] han encontrado metodologías las cuales se utilizan en este documento tal como se muestra posteriormente.

Una zonificación adecuada del circuito en el cual se harán las pruebas hace que la máquina funcione mejor. Se implementa en el circuito IEEE 34 nodos el cual se muestra en Figura 1. Con su respectiva zonificación (área donde se asumen los fallos).

Los parámetros de penalización se determinaron por pruebas realizadas anteriormente, en este documento se asumirán valores constantes.

La normalización de los datos de entrada es una función primordial, dado que grandes valores de entrada genera que el kernel no separe adecuadamente las diferentes muestras lo cual conlleva un lento entrenamiento y una mala clasificación de las zonas de falla, la normalización que ha presentado resultados buenos en la literatura

**Figura 1:** Zonificación circuito IEEE 34 nodos

Fuente.[Gil, 2011]

es la llamada normalización Min-Max y es descripta por la Ecuación (2).

$$S'_{ij} = \frac{s_{ij} - \min_j}{\max_j - \min_j} \quad (2)$$

Donde, S'_{ij} Es el conjunto de datos normalizados.

Es S_{ij} el conjunto de datos a normalizar.

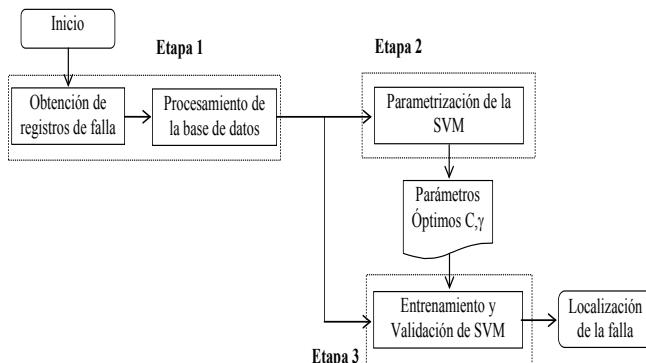
Es \min_j el valor mínimo del conjunto de datos j

Es \max_j el valor máximo del conjunto de datos j .

Esta normalización tiene buenos resultados gracias a que los datos se distribuyen en el intervalo $\{0-1\}$, diferentes normalizaciones se pueden ver en [Anil, 2005].

Uno de los principales parámetros que afectan en el desempeño del entrenamiento es el tipo de función a utilizar en el kernel para este documento se usa la función de base radial mostrada en la Ecuación (1)

La metodología básica aplicada al entrenamiento de máquinas de soporte vectorial es como se muestra en la Figura 2.

**Figura 2:** Esquema general metodológico para entrenamiento y validación.

Fuente. [Gil, 2011]

El enfoque general consiste en la misma metodología básica

Figura 2., pero el objetivo de este documento no es determinar unos parámetros C y γ (se asumen constantes en este documento), sino la combinación de descriptores adecuados para generar una localización de las fallas satisfactoria.

En este documento se presenta una búsqueda guiada en la combinación de variables de entrada a la máquina de soporte vectorial (descriptores del circuito), con el fin de minimizar un error de validación. La búsqueda consiste en la prueba del problema de clasificación teniendo como entradas la combinación exhaustiva, en grupo de dos, tres y cuatro descriptores.

Algunos de los descriptores se presentan en la Tabla 1, los cuales podrán ser relacionados con los resultados obtenidos con esta técnica de búsqueda.

Abreviatura	Descriptor
‘_dI’	Variaciones de las corrientes de fase entre prefalla y falla
‘_dV’	Variaciones de las tensiones de fase entre prefalla y falla
‘_dIL’	Variaciones de las corrientes de línea entre prefalla y falla
‘_dVL’	Variaciones de las tensiones de línea entre prefalla y falla
‘_dAngIlinea’	Variaciones de ángulo de corriente línea
‘_dAngVlinea’	Variaciones de ángulo de tensión línea
‘_dAngVfase’	Variaciones de ángulo de tensión fase
‘_dAngIfase’	Variaciones de ángulo de corriente fase
‘_Alf’	Angulo de la corriente de línea en estado de falla
‘_MIf’	Magnitud de la corriente de línea en estado de falla
‘_MVF’	Magnitud de la tensión de fase en estado de falla
‘_MVP’	Magnitud de la tensión de fase en estado de prefalla
‘_dFp’	Variaciones en el factor de potencia de fase
‘_dFpL’	Variaciones en el factor de potencia de línea
‘_dSL’	Variación en la potencia aparente total, entre prefalla y falla
‘_dX’	Variación en la reactancia de fase, entre prefalla y falla

Tabla 1: Descriptores del circuito

Fuente. Elaboración propia

3. RESULTADOS

La Tabla 2. resume las condiciones en las cuales se hacen las pruebas para el circuito IEEE 34 nodos.

Parámetro	Función
Kernel	Función de base radial
Normalización	Min-Max
Estado del circuito	Condición nominal
Zonificación	Predeterminada
Parámetros de penalización	$C=201656.7$ $=0.3316$
Número de fallas	1000
Número de fallas a validar	1000
Descriptores	Por determinar

Tabla 2:Condiciones de prueba

Fuente. Elaboración propia

Los resultados presentados a continuación fueron desarrollados en Matlab, y utilizado una máquina de soporte vectorial de uso libre.

Los descriptores a determinar en este documento fueron validados en una búsqueda exhaustiva desde la combinación de dos parámetros hasta la combinación de 4 de ellos (etapa 2 en la Figura 2). Las pruebas realizadas para combinaciones superiores dan resultados de mala calidad y entrenamiento sumamente lento, por lo tanto no se consideran en este artículo.

La prueba para las dos posibles combinaciones de descriptor es se muestra en la Figura 3. Cada descriptor es de un sistema trifásico real, entonces está conformado por un vector de tres elementos, por lo tanto en esta prueba el número real de descriptor es de seis. La combinación de dos descriptor es del circuito para la validación, genera 231 combinaciones, las cuales la máquina de soporte obtiene modelos aproximados y cada uno fue validado con una base de datos diferente a la del entrenamiento.

La Figura 3 muestra una señal que relaciona el número de combinaciones con respecto al error de validación, por lo tanto los picos más cercanos al cero son los valores de descriptor es que queremos utilizar para entrenamiento y validación; dado que estos conllevan un error menor de estimación en el número de fallas.

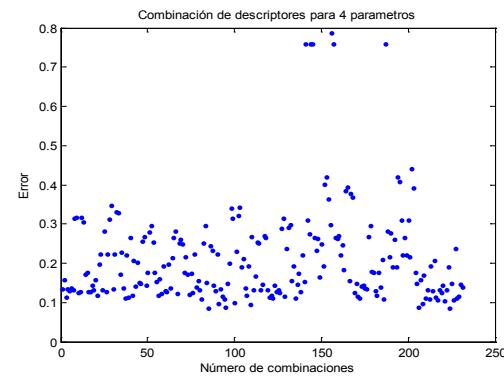


Figura 3: Combinaciones de dos descriptor es Vs error de validación

Fuente. Elaboración propia

En la Tabla 3 se muestran los 5 menores picos de la Figura.3 los cuales son las combinaciones de descriptor es adecuados para este procedimiento.

Número de validación	Error (%)	Descriptor es
224	8,44	'_Alf' '_dAngVlinea'
85	8,44	'_dI' '_dFpL'
94	8,67	'_dI' '_dAngIlinea'
204	8,67	'_MVF' '_Alf'
107	9,56	'_dIL' '_dAngIfase'

Tabla 3: Errores mínimos de Figura 3

Fuente. Elaboración propia

De la Tabla 3 se puede notar los diferentes descriptor es que dan un error de validación pequeño.

La Figura 4 es la gráfica que relaciona las diferentes combinaciones de descriptor es con el error de validación. Como las combinaciones de descriptor es son de tres elementos, entonces el número de combinaciones totales aumenta drásticamente para este caso el número de combinaciones es igual a 1540.

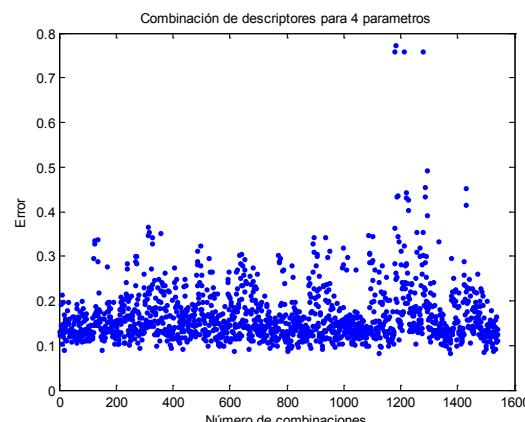


Figura 4: Combinaciones de tres descriptor es Vs error de validación

Fuente. Elaboración propia

En la Tabla 4 se hace la clasificación de los 5 menores picos de la Figura 4, estos son los descriptores que mejoran el desempeño de la máquina.

Número de validación	Error (%)	Descriptores
1121	8,22	'_dXL' '_dFp' '_dAngIlinea'
1373	8,22	'_dFp' '_dAngIfase' '_dAngIlinea'
1489	8,44	'_MVf' '_MIf' '_AIf'
615	8,67	'_dVL' '_dX' '_AIf'
800	8,67	'_dI' '_dFp' '_AIf'

Tabla 4: Errores mínimos de Figura 4

Fuente. Elaboración propia

De la búsqueda para 3 descriptores se puede ver que parámetros dan mejores resultados, respecto a la validación y entrenamiento Tabla 3.

En la Figura 5 se hace la combinación para 4 descriptores los cuales se validarán en la búsqueda exhaustiva y se encontrará la configuración óptima para los valores fijados los cuales son de mucha importancia en este proceso y en el error Tabla 2, el número total de combinaciones posibles para este caso fue de 7315.

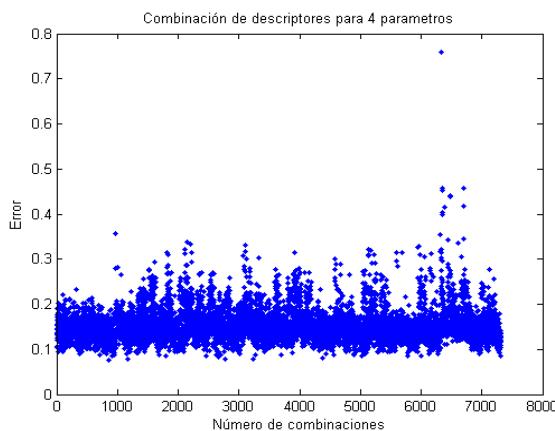


Figura 5: Combinaciones de cuatro descriptores Vs error de validación

Fuente. Elaboración propia

En la Tabla 5 se clasifican los 5 menores errores de validación, para encontrar los mejores parámetros.

Número de validación	Error (%)	Descriptores
2450	8,44	'_dSL' '_MIf' '_AIf' '_dAngIlinea'
5762	8,44	'_dX' '_dFp' '_MIf' '_dAngVfase'
5992	8,44	'_dXL' '_MVp' '_MVf' '_AIf'
7302	8,44	'_dAngVfase' '_dAngIfase' '_dAngVlinea' '_dAngIlinea'
713	8,67	'_dS' '_dIL' '_dFp' '_dAngIfase'

Tabla 5: Errores mínimos de Figura 5.

Fuente. Elaboración propia

De la Tabla 5 se puede notar la gran variedad de descriptores que pueden ser usado con los mismos resultados finales en error de estimación: es importante también recordar que cada uno de estos descriptores ya mencionados llevan 3 mediciones las cuales entran en la máquina de soporte como un conjunto que conforma el descriptor, entonces de este último caso en la máquina para cada validación y entrenamiento recibía 12 parámetros por cada combinación.

La mejor combinación de descriptores en el entrenamiento de la máquina fue de tres parámetros y fueron '_dXL' '_dFp' '_dAngIlinea', estos presentaron un error de validación menor, también los parámetros '_dFp' '_dAngIfase' '_dAngIlinea' obtuvieron el error de validación más bajo, se puede notar que un descriptor se encuentra en los dos grupos, este parámetro presenta datos altamente separables por la metodología de entrenamiento de la máquina, para ello se debe hacer un análisis más exhaustivo pero se puede decir que utilizar tres descriptores es más eficiente.

4. CONCLUSIONES

A partir de los resultados obtenidos se puede determinar que los errores de validación son altamente dependientes de la calidad de los descriptores utilizados, dado que una variación en ellos puede hacer aumentar o disminuir el error de localización de fallas.

Adicionalmente, es importante notar que todos los resultados están relativamente cercanos en precisión del localizador de fallas, ya que las mejores combinaciones

de dos parámetros es tan buena como la combinación de cuatro parámetros, entonces es importante escoger el menor número de parámetros, dado que el tiempo computacional al tener menos descriptores se reduce considerablemente.

Como recomendación importante, se determinó que al encontrar los descriptores de mejor calidad se puede correr nuevamente el proceso mostrado en Figura 2, para hallar nuevamente los parámetros de penalización que reducen al mínimo los errores de validación cruzada.

Finalmente, se debe considerar que cada circuito tiene condiciones diferentes de operación entonces, se debe analizar esta metodología con muchas más condiciones y así poder hacer una máquina más robusta. Esta máquina de localización ayuda a mejorar los índices de continuidad del suministro y así mejorar el funcionamiento del sistema de potencia y los clientes atendidos.

5. AGRADECIMIENTOS

Este trabajo se realizó en el Grupo de Investigación en Calidad de Energía Eléctrica y Estabilidad (ICE3) financiado por la Universidad Tecnológica de Pereira (UTP) y por medio del proyecto de investigación “Desarrollo de localizadores robustos de fallas paralelas de baja impedancia para sistemas de distribución de energía eléctrica -LOFADIS2012-”, contrato 0977-2012, financiado por Colciencias.

REFERENCIAS

Anil J. , Karthik N., Arun R., “Score normalization in multimodal biometric systems”, *Pattern Recognition*, vol. 38, no.12, pp. 2270–2285, Dec. 2005.

Gil, W.J. (2011). *Utilización de técnicas metaheurísticas en la búsqueda de parámetros óptimos para la calibración de las máquinas de soporte vectorial (SVM), para la localización de fallas en sistemas de distribución*. Tesis de pregrado, Programa de Ingeniería Eléctrica, Universidad tecnológica de Pereira, Pereira, Colombia

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*, 2010, no. 53, pp. 174-187. Jul. 2010.

Lawrence D.J., Cabeza L., and Hochberg L., “Development of an advanced transmission line fault Location system part II—algorithm development and simulation,” *IEEE Trans. Power Delivery*, Vol. 7, No. 4, Oct. 1992, pp. 1972–1983.

Maldonado S., “Utilización de support vector machines no lineal selección de atributos para credit scoring”, Tesis de Maestría, Universidad de Chile, Santiago de Chile, Chile, 2007.

Maldonado S., R. Weber, “Modelos de Selección de Atributos para Support Vector Machines”, *Revista Ingeniería de Sistemas*, 2012, no 26, pp. 49-70, Sep. 2012.

Moguerza J., Muñoz A., “Support Vector Machines with Applications”, *Statistical Science*, vol. 21, no.4, pp. 322-336, 2006.

Mora J., Carrillo G., Meléndez J., “Comparison of impedance based fault location methods for power distribution systems”, *Electric Power Systems Research*, vol. 28, no. 7, pp. 657–666, Apr. 2008.

Mora, J.J. (2006). *Localización de Faltas en Sistemas de Distribución de Energía Eléctrica usando Métodos basados en el Modelo y Métodos de Clasificación Basados en el Conocimiento*. Tesis Doctoral. Departamento Tecnologías de la Informática, Universidad de Girona, Girona, España.

Morales G., Mora J., 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, pp. 803-809, Nov. 2009.

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, pp. 710-721, Apr. 2005.