Evaluation of two knowledge-based fault locators for power distribution systems

Evaluación de dos localizadores de fallas basados en el conocimiento para sistemas de distribución de energía eléctrica

N. Alzate-González[1], A. Zapata-Tapasco[2] , J. Mora-Flórez [3]

ABSTRACT

This paper presents the implementation of two knowledge-based methods to locate faults on power distribution systems with distributed generation, to improve the power quality as demanded by the regulated electrical markets. Fault location methods could use additional measurements as new information to improve the results. ID3 decision tree and *k* nearest neighbor methods are implemented using measurements at substation and at the distributed generation nodes. The methods are validated using a modification of the IEEE 34-node test feeder, with two distributed generators. Several operating scenarios are considered by varying the model parameters of the power system. Additionally, a sensitivity analysis is performed to find the most critical parameters that affect the fault location methods. The results show the good performance of the fault locators and help to determine the most critical parameters of the power distribution system.

Keywords: Distributed generation, fault location, knowledge-based methods, sensitivity analysis.

RESUMEN

En este artículo se presenta la implementación de dos métodos de localización de fallas basados en el conocimiento para sistema de distribución de energía considerando generación distribuida, para mejor la calidad de la energía según lo exigido por la regulación de los mercados eléctricos. Los métodos de localización de fallas pueden utilizar las medidas adicionales como una nueva información para mejorar los resultados. Se implementa el método de árbol de decisión ID3 y los *k* vecinos más cercanos usando medidas en la subestación y en los nodos con generación distribuida. Los métodos se validaron utilizando una modificación del alimentador de prueba IEEE-34 nodos, con dos generados distribuidos. Se consideran varios estados operativos variando los parámetros de modelado del sistema de potencia. Adicionalmente, se realiza un análisis de sensibilidad para encontrar los parámetros más críticos que afectan los métodos de localización. Los resultados muestran un buen desempeño de los localizadores de fallas y ayudan a determinar los parámetros más críticos del sistema de distribución de energía.

Palabras clave: Generación distribuida, localización de fallas, métodos basados en el conocimiento, análisis de sensibilidad.

Received: July 24th 2015
Accepted: September 15th 2015

Introduction

As a consequence of the privatization of the electricity sector, the continuity of energy service is nowadays a topic of great interest for the utilities, which must pay penalties in the case of not meet the quality levels established by the regulatory agencies [Salim, et al., 2011].

Nowadays by using the new technological developments, many power distribution systems have implemented generation centers based on alternative energy sources as solar, wind, among others. Distributed generation systems (DG) introduce new problems in the control and protection of the power system due the bidirectional currents, which affects power quality [Orozco, et al., 2012].

Some fault location methods have been implemented to help in the solution of supply continuity problem, by reducing the outages duration, having a fundamental role in the fast and reliable process of service restoration [Mora, et al., 2006]. However, most of these methods were developed based on the radial nature of the power system.

On the other side, several fault location methods have been developed to determine a faulted zone based on data mining methods, and are called knowledge-based methods (KBM). Some KBM successfully implemented for fault location problem are *k* nearest neighbors (*knn*) [Mora, et al., 2009] and decision trees (ID3) [Zapata, et al., 2014].

[1] N. Alzate-González, Electrical engineer Universidad Tecnologica de Pereira, Colombia. Affiliation: Researcher at ICE3, Universidad Tecnológica de Pereira, Colombia. E-mail: naalzate@utp.edu.co.

[2] A. Zapata-Tapasco: Electrical engineer, M.Sc. Universidad Tecnologica de Pereira, Colombia. Affiliation: Researcher at ICE3, Universidad Tecnológica de Pereira, Colombia. E-mail: anfezapata@utp.edu.co.

[3] J. Mora-Florez: Electrical engineer, M.Sc Universidad Industrial de Santander, Bucaramanga, Colombia. Ph.D., Universitat de Girona, España. Affiliation: Associated professor at Universidad Tecnológica de Pereira, Pereira, Colombia. E-mail: jjmora@utp.edu.co

How to cite: Journal Space xxxxxxxxxxxxxxxxxxxxx
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

KBM are not based on the system model, since the zone under fault is determined using the measurements of voltage and current at the generation sources. Because of this, KBM can improve the results previously obtained for power distribution systems by using the measurements at the DGs. This research focuses on KBM applied to power distribution system that contains distributed generation systems.

Using the sensitivity analysis evidences the KBM dependence on the model parameters of the power distribution system. This dependence is a complex problem, because in power systems, many parameters have a considerable degree of uncertainty, especially those related to the load.

This paper presents the implementation of two knowledge-based methods applied to electric power distribution systems, which consider the presence of distributed generation. Additionally, a sensitivity analysis is performed to determine the model parameters of the power distribution system that significantly contribute on the performance of the fault locator. Therefore, this approach helps to reduce dependence on modeling parameters, which allows the development of more robust locators.

The paper is structured as follows; section 2 presents the theoretical aspects of the modern power distribution systems, knowledge-based fault locators and the sensitivity analysis. In section 3, the proposed methodology to implement the fault locator is described. The results of two KBM are discussed in section 4 and the conclusions are presented in section 5.

Theoretical Aspects

A. *Modern power distribution systems*

At modern power distribution systems, the distributed generation has an increasing importance due to government public policies and regulations and also due the development of many technologies, allowing large-scale implementation within the existing power electric system [Puttgen, et al., 2003]. Its implementation and integration into an existing utility can result in several benefits including line loss reduction, reduced environmental impacts, peak shaving, increased overall energy efficiency, relieved transmission and distribution congestion, voltage support, and deferred investments to upgrade existing generation, transmission, and distribution systems [Chiradeja, 2005].

However, it also implies several problems that must be take into account such the presence of harmonics due to converters, coordination of reactive power (specially with wind turbines), and the need of design and implement a more sophisticated protective relaying schemes, that considers bidirectional currents and provide safety to the personnel working on the lines [Puttgen, et al., 2003].

B. *Fault locators based on knowledge-based methods KBM*

KBM are based on data mining, which has become increasingly important due to the amount of data that is registered in real life processes. This dataset contains hidden information that can be seen using computational tools. One task of data mining is the classification, which aims to assign each input set of characteristics to one of a discrete set of categories or classes [Bishop, 2006]. Classification techniques belong to supervised learning, which means that each training data has a label that identifies it in one of the categories (classes) [Moguerza & Muñoz, 2006] [Burges, 1998].

Some classification techniques successfully used to solve the fault location problem are:

1. *K nearest neighbor*

This algorithm considers that an input vector belongs to the most frequent class, between its *k* nearest neighbors [Cover & Hart, 1967]. As advantage, this method has an easy implementation process and also has well defined statistical characteristics, which leads it as a very attractive technique to be used in any classification problem [Moreno, 2004].

2. *Decision trees*

A decision tree is a classification model, which divides the input space into cuboid regions, whose edges are aligned with the axes [Bishop, 2006]. Each region represents a category or class. Decision trees are based on the divide and conquer idea, developed by Hunt [Quinlan, 1993].

C. *Sensitivity analysis*

Sensitivity analysis studies the relation between the input and the output variable of a model. There are three kinds of methods for sensitivity analysis: selection, local and global [Tarantola, et al, 2004].

Selection methods identify a set of input variables affecting the output. These methods provide a qualitative measure and use little computational effort. Local methods identify the input variables that affect the output, but making very small changes in the input variables. Commonly, local methods vary one variable at a time while the others remain constant. Finally, global methods determine the influence of input variables in the model output. In this method, the input variables vary within different uncertainty ranges [Saltelli & Chan, 2000].

To perform sensitivity analysis is necessary to identify the problem or weakness of the model and determine the input variables that must be considered in the analysis. An uncertainty range is assigned to each input variable, according to the problem being analyzed.

Subsequently, the sensitivity analysis generates several scenarios to study the behavior of the model. A scenario is obtained assigning to each input variable a value within its uncertainty range. However, there are many combinations of values for the input variables and evaluate all of these implies a very high computational cost. Therefore, to reduce complexity, a sample of the total population is obtained through a sampling technique.

Latin hypercube sampling technique is used, which generates an *n* by *s* uncertainty matrix, where *n* represents the number of scenarios to assess and *s* represents the number of input variables to analyze. The uncertainty matrix generates values between zero and one, indicating the percentage change of each variable in each scenario [Viana & Venter, 2009] [Liefvendahl & Stocki, 2005].

Then, the model is executed repeatedly to obtain the output variable for each scenario. Finally, a statistical technique is selected for assessing the importance of each input variables with respect to the uncertainty in the output. In the literature, there are several statistical techniques to analyze the relation between the input and output variables of a model. Some of the most used techniques are: the analysis of variance (ANOVA), correlation coefficients and regression analysis.

Proposed Methodology

The proposed methodology for KBM implementation for a distributed generation system is explained below. ATP is used to model the power system and MATLAB is used for data processing.

The proposed methodology consists of four steps to find the performance of the KBM method. The last step is aiming to find the model parameters that have the bigger dependency on the performance.

A. Stage 1. Data acquisition

1. Model parameters

In this paper and due to the nature of the problem, the global sensitivity analysis was chosen. The model studied in the global sensitivity analysis is a KBM, which uses voltage and current measurements at the generation sources to calculate the faulted area. However, these measures depend on the power system model; therefore, the input variables to be analyzed are the model parameters.

To test the confidence of the proposed KBM and to simulate more accurately the distributed power system, the model parameters vary within a corresponding uncertainty range. Variation of the model parameters is performed to consider different operating states of the power system, which allows the analysis of the fault location methods in case of uncertainties at the input variables. The modeling parameters are: power factor, voltage source magnitude, voltage source unbalance, soil resistivity, system load magnitude and line length.

2. Create several operating scenarios

Using a sampling technique, *n* operating scenarios of the power system are created, each one with a different parameter variation, to study the behavior of the model in these scenarios. Latin hypercube is used, and a computational tool is implemented to vary the rated system modeled in ATP.

The computational tool automatically varies the model parameters of the power system using as input data the uncertainties matrix created by the latin hypercube and the rated system modelled in ATP. Every row of the matrix shows the percentage change in each model parameter. With these percentages determines the new value of the parameters and replaces these on the rated system, creating all operating states used to evaluate fault location methods.

3. Obtain the fault database

After being modeled all operating scenarios in ATP, a computation tool automatically simulates faults [Alzate, et al., 2014]. This tool adds a fault element to each operation state according to fault type and their fault resistance. ATP simulation provides voltage and currents signals for each generation source. These signals are converted to MATLAB for the data processing. Finally, a computation tool obtains the voltage and current phasor values, at pre-fault and fault stage in each generation source, which are the input data to evaluate the fault location method.

B. Stage 2. Zone definition and data preprocessing

The power distribution system is subdivided into zones, which are the target of the classification method. This subdivision must takes account several aspects such as the circuit topology, zone size, location of protective devices and availability of enough data in each zone to train the classification tool [Mora, et al., 2009]. In this paper, the automatic zone definition proposed in [Zapata, 2013] was used.

Data preprocessing consist in to label each fault with the respective zone. Others task as outliers detection, feature selection, noise cancelation and strategies to handle missing data are not performed.

C. Stage 3. Evaluate KBM

ID3 decision tree and *k* nearest neighbor are validated using 10-fold cross validation error. *Knn* optimal parameters are found using a fault database at rated condition. The performance index is obtained for every operating condition using the two KBM fault locators. The overall cross validation error obtained measures the confidence of the KBM based fault locators.

D. Stage 4. Evaluate a sensitivity technique

Regression analysis is used to indicate the importance of each input parameter with respect to the uncertainty in the output. Regression analysis is performed using the least squares method, which is explained in [Saltelli & Chan, 2000].

In this paper, least squares regression analysis determines the importance of the model parameters in fault location. This technique has as input data, the uncertainties matrix obtained from the latin hypercube, and the performance indices obtained by the fault location method for each operating condition. Finally, standardized Beta coefficients for each parameter are obtained. These coefficients have values between zero and one, indicating the influence of the parameters variation on the KBM fault locators.

Results and Discussion

Two knowledge-based methods were validated using the IEEE 34-node system, which is taken from the "Distribution System Analysis Subcommittee" [Radial Test Feeder]. The power systems were modified by adding two distributed generation sources at nodes 824 and 836. Figure 1 shows the power system and the zone definition.

800 operating states were generated by varying six model parameters simultaneously such as power factor, voltage source magnitude, voltage source unbalance, soil resistivity, system load magnitude and line length. The uncertainty range of each parameter is shown in Table 1. In addition, fault resistance varies from 0.05 to 40 Ω [Dagenhart, 2000].

Load magnitude has a larger variation because of the uncertainty of this parameter on the distribution system; this variation can be seen on the 'daily load profile for residential consumers'.

The variation range of the power factor, voltage source magnitude and unbalance is set according to the current regulations in Colombia. Soil resistivity variation is set according to soil characteristics, taking into account the composition, temperature and moisture of soil [Mora, et al., 2010]. And the line length variation is taken account because of possible errors in the data base of the utilities or considering variation due to temperature changes, the time and the voltage.

Single-phase A-G fault was simulated, obtaining 129600 faults at different nodes of the power system.

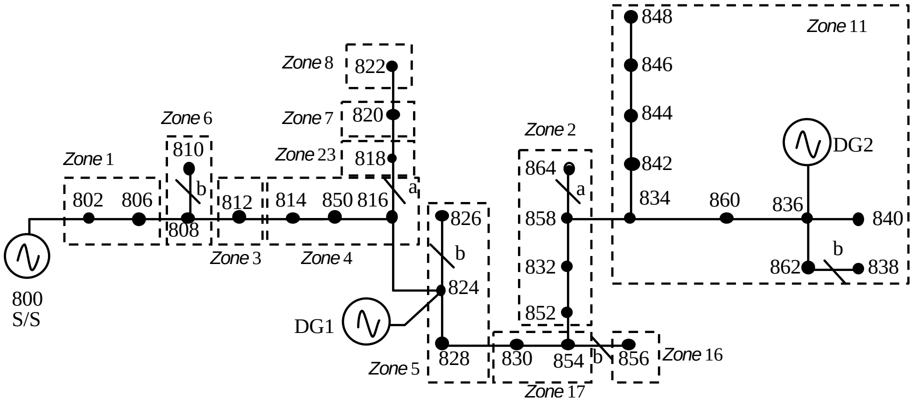

Fig 1. IEEE 34 test feeder. “Own compilation”.

For *knn* method, an overall cross validation error of 8.81% was obtained for single-phase faults. For ID3 decision tree method, an error of 1.32% was obtained for single-phase faults. Both methods have a small overall cross validation error, which is attributable to the inclusion of distributed generation. Fault location methods use measures of distributed generators to improve the results and eliminate errors associated with the uncertainty in the model. The performance of *knn* method can be improved using a larger fault database or using a cross-validation with more folds.

The results of the sensitivity analysis methodology for knn and ID3 decision tree are shown in Fig. 2 and 3. Figure 2 shows the Beta coefficients obtained with *k* nearest neighbor method for single-phase A-G faults. It can be seen that the parameter that most affects the locator performance is system load magnitude.

On the other hand, a variation of power factor does not affect the fault location performance. Therefore, this parameter is insignificant for the *k* nearest neighbor method.

As shown in Fig. 3, for single-phase A-G, the parameter that most affects the ID3 fault locator performance is system load magnitude; however, in this case, beta coefficients of these parameters only reach up to 0.19, which indicates a low dependency of the parameters on the performance.

The results of both methods show that system load magnitude has the greater influence on the performance. This was expected because this is the parameter with a larger variation.

Table 1: Uncertainty ranges of the parameters. “Own compilation”.

Model parameters	**Uncertainty range**	
	Minimum	**Maximum**
Power factor	-0.02	0.02
Voltage source magnitude	0.95 p.u	1.10 p.u
Voltage source unbalance	-3.4°	3.4°
Soil resistivity	80 [Ω*m]	120 [Ω*m]
System load magnitude	10%	150%
Line length	95 %	105 %

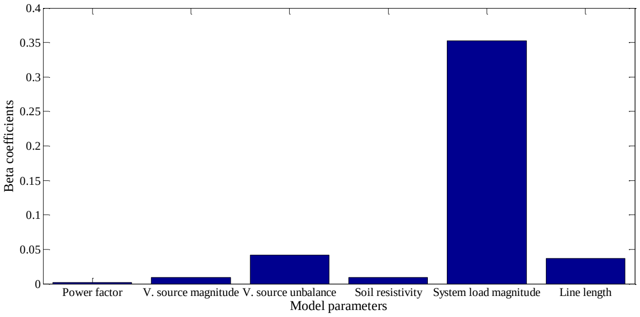

Fig. 2. Results of the sensitivity analysis with k nearest neighbor method. “Own compilation”.

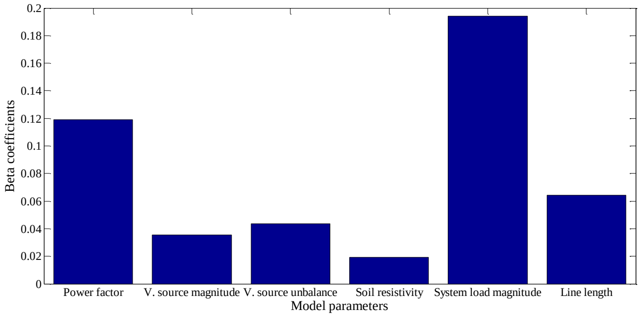

Fig. 3. Results of the sensitivity analysis with ID3 decision tree method. “Own compilation”.

Concluding remarks

This paper presents an approach to locate the faulted zone on power distribution system with distributed generation using two knowledge-based methods. The 10-fold cross validation error was used to find the confidence of the methods.

ID3 decision tree and *k* nearest neighbor demonstrated a very good performance on the fault location problem at power distribution systems with GD, which is seen on the performance of ID3 with an average error of 1,32% and *knn* with an average error of 8,81%.

Finally, a sensitivity analysis for ID3 decision tree and *k* nearest neighbor was performed. The obtained Betas coefficients for both methods are small, especially for the ID3 method, which indicates that this method have a low dependence on the model parameters. A further analysis of *knn* method using a different cross-validation has to be performed to confirm the obtained results.

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 of Universidad Tecnológica de Pereira (Colombia).

References

Alzate, N., Mora, J., Pérez, S., Methodology and software for sensitivity analysis of fault locators, Transmission & Distribution Conference and Exposition-Latin America (PES T&D-LA), 2014 IEEE PES, Sept. 2014, pp. 1-5.

Bishop, C.M., Pattern Recognition and Machine Learning, New York, Springer-Verlag, 2006.

Burges, C., A tutorial on support vector machines for pattern recognition., Data Mining and Knowledge Discovery, 1998, pp. 121-127.

Chiradeja, P., Benefit of Distributed Generation: A Line Loss Reduction Analysis, Transmission and Distribution Conference and Exhibition: Asia and Pacific, 2005 IEEE/PES, 2005, pp. 1-5.

Cover, T., Hart, P., Nearest neighbor pattern classification, Information Theory, IEEE Transactions on, Vol. 13, No. 1, 1967, pp. 21-27.

Dagenhart, J., The 40-Ω ground-fault phenomenon, Industry Applications, IEEE Transactions, Vol. 36, No.1, Feb. 2000, pp. 30-32.

Liefvendahl, M., Stocki, R., A study on algorithms for optimization of Latin hypercubes, Journal of Statistical Planning and Inference, 2005, pp. 3231-3247.

Moguerza, J.M., Muñoz, A., Support Vector Machines with Applications. Statistical Science., 2006, pp. 322-336.

Mora, J., Melendez, J., Bedoya, J., Extensive Events Database Development using ATP and Matlab to Fault Location in Power distribution Systems, IEEE PES Transmission and Distribution Conference and Exposition: Latin America, Caracas, 2006.

Mora, J., Morales, G., Pérez, S., Learning-based strategy for reducing the multiple estimation problem of fault zone location in radial power systems, Generation, Transmission & Distribution, IET, Vol. 3, No. 4, April 2009, pp. 346-356.

Mora, J., García, G., Pérez, S., Fault resistance impedance based methods for locating faults. A comparative analysis, Dyna, Universidad Nacional de Colombia, 2010.

Moreno, F., Clasificadores eficaces basados en algoritmos rápidos de búsqueda del vecino más cercano, Ph.D. dissertation, Universidad de Alicante, Departamento de lenguajes y sistemas informáticos, España, 2004.

Orozco, C., Mora, J., Pérez, S., A robust method for single phase fault location considering distributed generation and current compensation, Transmission & Distribution Latin America Conference, Montevideo-Uruguay, Sept. 2012.

Puttgen, H.B., MacGregor, P.R., Lambert, F.C., Distributed generation: Semantic hype or the dawn of a new era?, Power and Energy Magazine, IEEE, Vol.1, No.1, Feb. 2003, pp. 22-29.

Quinlan, J., Programs for Machine Learning, Morgan Kaufmann Publishers, San Mateo California, 1993.

Radial Test Feeder, IEEE Distribution System Analysis Subcommittee. Available: http://ewh.ieee.org/soc/pes/dsacom/testfeeders/index.html.

Salim, R.H., Salim, K.C., 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.

Saltelli, A., Chan, K., Sensitivity Analysis, John Wiley & Sons Ltd, United Kingdom, 2000, pp. 101-152.

Tarantola, S., Saltelli, A., Campolongo, F., Ratto, M., Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models, John Wiley & Sons Ltd, United Kingdom, 2004.

Viana, F., Venter, G., An algorithm for fast optimal Latin hypercube design of experiments, International journal for numerical methods in engineering, United States, 2009.

Zapata, A., Implementación y comparación de técnicas de localización de fallas en sistemas de distribución basadas en minería de datos, Tesis de Maestría, Departamento de Ingeniería Eléctrica, Universidad Tecnológica de Pereira, 2013.

Zapata, A., Mora, J., Cortes, M., Fault location in power distribution systems using a learning approach based on decision trees, Transmission & Distribution Conference and Exposition - Latin America (PES T&D-LA), 2014 IEEE PES, Sep. 2014, pp. 1-6.