A multi-stage methodology for fault location in radial distribution systems

André Luís da Silva Pessoa
Mário Oleskovicz
Paulo Estevão Teixeira Martins
Department of Electrical and Computer Engineering
University of São Paulo, USP
São Carlos, Brazil
alsp@usp.br, olesk@sc.usp.br, pauloetm@usp.br

Abstract—In the context of smart grids and microgrids, fault location is an important point that can be explored with all the possibilities of this new paradigm. This study presents a multi-stage methodology that involves decision trees as well as artificial neural networks for fault location purposes. Another aspect to be considered is the use of an allocation of smart meters in the distribution system for the tests. The IEEE 34-bus distribution system was used to validate the methodology. This system was chosen due to its considerable number of monophasic branches and its extension. The focus of this paper is the multiple estimation problem, in which, for a same distance of the fault, there is more than one possibility to locate it. Confusion matrices were used to explore the results and show the performance of the proposed algorithm.

Index Terms—Artificial Neural Network. Decision tree. Fault location. Smart Grid. Smart meter allocation.

I. INTRODUCTION

The current changes in the system due to, for example, the introduction of distributed generation and electric vehicles, has brought new challenges to the traditional Electrical Power System (EPS) model. These changes in the electric scenario occurs due to the need of more sustainable sources, more efficient transportation (and final use) of electric energy, and also to modernize the infrastructure of the electric sector [1]. To prevent system failures, some countries have been investing heavily in renewing power grids. This increases the interest in improving the current electrical networks by incorporating intelligence for each small component of the system, as well as in the design of the electrical system as a whole [2]. Thus, ensuring that there is no loss in the quality of the service provided by the electric utilities to the final users (consumers).

Smart grids, in general, positively affect the following aspects associated with electrical power systems: reliability; flexibility; efficiency; sustainability; degree of intelligence in management/operation; and the increasing use of technologies that enable real-time communication between generation, transmission, distribution and consumers. [3], [4] comment that the number of sensors in smart grids, as well as the processing capacity, will increase when compared to the number of sensors present in current electrical systems. In this scenario, and for example through wireless communication, smart meters can provide the necessary information of all points of interest in a given electrical system to a master element, which can then establish real-time control of the system as a whole [4].

Thus, taking into account that fault situations are unpredictable events occurring in the system [5], and that a fast location of these disturbances is an essential element for the restoration of an acceptable power quality for users affected by these phenomena, the whole infrastructure present in smart grids should be used to ensure a better quality of the service and of the product to system users.

In the various papers related to fault location in distribution systems (that use some artificial intelligence approach), it is observed that variations, such as the fault type (single-phase to ground, double-phase, double-phase to ground, three-phase and three-phase to ground), fault impedance, fault incidence angle, and fault position in the test system, were considered to form the database for the training of the learning algorithms used. To form the representative dataset, in general, two types of approach were observed: one in which the validation database has the same profile as the database of the training phase; and cases where the validation database has a distinct nature from the data set used for the training process. Another issue to be highlighted is the number of measurement points and whether or not meters optimally positioned in the system [6]–[12]. [6] used a clustering algorithm to locate faults based on values of voltage sags observed in the system. [7] used Support Vector Machines (SVM) for the classification of short-circuits types and Artificial Neural Networks (ANNs) to estimate some regions in which the fault could have occurred. [8], [12] used ANNs to estimate the distance of the fault. [9] used the algorithm radial basis function neural network with optimum steepest descent to locate faults in a distribution system. In the proposed methodology, data from various points of the system were used for the correct location of the fault. [10] used a decision tree-based methodology to locate the regions of the system in which the fault occurred. [11] used the Wavelet Packet Transform to extract parameters from the post-fault signals and the Radial Basis Function
Neural Network to estimate the distance of the fault.

As shown, the various researches on fault location either estimate the distance from the fault or identify the region (branch) of the system in which the fault occurred. In order to better deal with the problem of multiple estimation, this paper proposes the use of a multi-stage methodology using artificial neural networks and decision trees for the correct location of faults in a radial distribution system.

In addition to this introduction, this article is divided into 4 more sections. Section 2 presents some points related to artificial neural network algorithms and decision trees adopted in this paper. Section 3 presents the methodological aspects considered. Section 4 contains some results. Section 5 presents the conclusions.

II. ARTIFICIAL NEURAL NETWORK AND DECISION TREES

Artificial neural network algorithms are bioinspired in biological neural networks. ANNs of the Multi-layer Perceptron type have supervised training. Therefore, it is necessary to have input patterns and their corresponding desired outputs. Thus, ANNs are able to create an internal model that learns about the behavior of a system [13]. The function approximation is one of the different uses for this algorithm and was used in this article to estimate the distance of the faults in relation to the substation. Thus, the input parameters of the ANNs used are the RMS (Root Mean Square) values of the voltage and current signals observed in the substation and the output of the ANN is the estimation of the distance of the fault in relation to the substation.

Decision trees also have supervised training. Another point to be highlighted is the ability of decision trees to perform an internal attribute selection. Thus, at the end of its training, a node structure is created in which the intermediate nodes (leaf nodes) are the parameters chosen by the algorithm as the most significant, and the final nodes are the identified classes that are present in the output of the database [13], [14]. In this research, decision trees were used to identify the areas and regions in which the faults occurred.

Further details on the input parameters of these algorithms will be presented in the next section.

III. PROPOSED METHODOLOGY

Fig. 1 presents the multi-stage methodology adopted for the fault location. The first step represents the obtaining and processing of the signals of the second post-fault cycle of the three-phase voltage and current signals. After that, the obtained parameters are used to initially estimate the distance of the fault (by using an ANN with back-propagation training algorithm) in relation to the substation, and to identify the area of occurrence of the fault (more general location). With the parameters obtained by the processing of the three-phase voltage and current signals, by fault distance estimation and by the identification of the fault occurrence area, decision trees (with CART - Classification and Regression Trees - training algorithm) are used to identify the region (a more specific location) of the occurrence of the fault.

This section is divided into the following subsections: the system used for training and testing the proposed methodology; simulations performed to form a dataset; parameters extracted from the fault situations applied; definition adopted for area and region; and application of the fault location algorithm.

![Fig. 1: Methodology for fault location.](image-url)
a sampling rate of 256 samples/cycle. Three meters were empirically allocated for the application of this methodology in the system (according to Fig. 2). This system was chosen due to the fact that it has a considerable extension, it is unbalanced and presents single-phase lateral branches. It is worth noting that, for the fault simulations, only the medium voltage sections were considered.

In the context of this research, 114 parameters were calculated after a fault detection cycle. The calculated parameters were:

1. RMS value;
2. the amplitudes and phases of components of the fundamental frequency; and
3. the energy of the first two coefficients of the 4th level of Wavelet Packet Transform (WPT) decomposition [11], using Daubechies with support 4 (db4) as the mother wavelet.

In this paper, the distance from the fault in relation to the substation was estimated in each fault situation. The RMS value of three-phase voltage and current signals from the substation was used for the estimation of the fault distance. It is worth mentioning that only the RMS values of the voltage and current signals of the phases (or phase) directly involved in the fault situation were used. Thus, for example, the RMS values of the voltage and current of phase A were used for the single-phase A-to-ground faults. For the double-phase A-B-to-ground faults, for example, the RMS values of the voltage and current signals of the A and B phases were used as input parameters. Thus, for each type of fault, a single decision tree that will identify the area of occurrence of the fault was trained.

Regarding the identification of the region of occurrence of the fault, a decision tree was trained for the 114 data coming from each of the meters installed in the system. In addition, these trees have as input the estimation of the distance of the fault and the identification of the area of occurrence of the fault. Therefore, each of these trees has 116 input parameters. Another point is that to the end, as each tree will perform an identification of the region of occurrence of the fault, for the final answer, the redundancy of the responses of all these trees will be considered. That is, if all trees are unanimous in the response, the answer of all will be adopted, if there is no majority, the response of the tree that belongs to the area in which the fault occurred will be considered.

D. Area and region definitions

Fig. 3 and 4 present the area and region definitions adopted in this paper. The criterion used in [17] was adopted for the definition of area and region. Thus, 3 areas and 10 regions were defined.

E. Application of the fault location algorithm

In general, after detection and classification (single-phase to ground, double-phase, double-phase to ground, three-phase and three-phase to ground) of the fault, three-phase voltage and current signals of the substation and of the meters allocated to the system are used by the fault location algorithm. Initially, a processing of these signals is performed for the extraction of parameters that will characterize the fault behavior throughout the system. Then, the value of the signals coming from the substation are used to estimate the distance of the fault. In
parallel, all parameters coming from all the meters are used to identify the area of occurrence of the fault. After having both the estimation of the distance of the fault and the identification of the area, the identification of the region of occurrence of the fault is performed. At the end, the algorithm will provide the information of the estimated distance of the fault and the region of the system in which it occurred.

IV. RESULTS

This section presents the results of the proposed methodology for the location of single-phase faults (A-to-ground, B-to-ground, and C-to-ground) in IEEE 34-bus distribution system. Initially, the results will be presented concerning the fault distance estimation. Then, the results will be shown for the identification of the area and the region of short-circuits occurrence. For a better analysis, the results will be presented for each type of fault.

A. Results for fault location

The results obtained for the estimation of the faults distance in relation to the substation are shown in Table I. For each of the single-phase faults evaluated, the values were calculated considering the percentage error: i) the 25th percentile (P25); ii) the 75th percentile (P75); iii) the average; iv) the standard deviation; and v) the kurtosis.

Among the three types of short-circuits evaluated, phase C-to-ground fault presented the lowest value for P25 and a higher value for P75. This indicates that, for the estimation of the fault distance, when phase C-to-ground faults are evaluated, the limits of values in which a good part of their data fits, present both a lower value and a larger value for the percentage error.

The values of the mean error and standard deviation are both higher for single-phase C-to-ground faults. In addition to a mean error slightly higher, the mean error tends to vary more with respect to the average value.

Finally, the kurtosis value of the phase C-to-ground faults also presents a higher value in relation to the other three fault types evaluated. This coefficient reveals the presence of outliers in the data. In these cases, there are far more outliers in the observed percentage error values for the phase C-to-ground faults than for the phase A and B-to-ground faults.

B. Results for outage area location

For the outage area identification, a hit rate of 100% was obtained for each of the three fault types evaluated.

C. Results for outage region location

A total hit rate of 95.8% was obtained for the faults A-to-ground, 89.8% for the faults phase B-to-ground, and 92.4% for the faults phase C-to-ground. Therefore, it is evidenced that, although faults phase C-to-ground had a worse performance in the estimation of the distance of the fault, faults phase B-to-ground presented a lower hit rate in the identification of the region of occurrence of the fault.

In order to have a more accurate analysis of the performance of the algorithm, the confusion matrices were used. Fig. 5, 6 and 7 present, respectively, the confusion matrices for fault region identification considering phase A-to-ground, B-to-ground and C-to-ground faults.

By analyzing Fig. 5, it is observed that there were no errors in locating faults that occurred in regions 1, 3, and 4, and that there was a minimum hit rate of 95% in the location of faults which occurred in regions 7, 9 and 10. It is observed that it is more difficult to locate faults that occurred in region 5. Thus, of all faults that occurred in region 5, the fault localization algorithm correctly located 83 faults, and incorrectly determined that 60 faults occurred in region 3.

In the confusion matrix of Fig. 6 region 1 was the only region in which the algorithm had a 100% of hit rate. The
TABLE I: Results for fault location.

<table>
<thead>
<tr>
<th>Fault type</th>
<th>P25 (%)</th>
<th>P75 (%)</th>
<th>Mean error (%)</th>
<th>Standard deviation (%)</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>phase A-to-ground</td>
<td>1.17</td>
<td>4.56</td>
<td>3.68</td>
<td>3.36</td>
<td>5.68</td>
</tr>
<tr>
<td>phase B-to-ground</td>
<td>1.0</td>
<td>5.34</td>
<td>3.28</td>
<td>2.73</td>
<td>3.88</td>
</tr>
<tr>
<td>phase C-to-ground</td>
<td>0.58</td>
<td>5.92</td>
<td>3.97</td>
<td>5.32</td>
<td>22.99</td>
</tr>
</tbody>
</table>

Algorithm had a hit rate above 85% in the fault location which occurred in regions 3, 7, 8, 9 and 10. It was more difficult to locate the faults that occurred in regions 2, 5 and 6. For the faults that occurred in region 2, the algorithm considered that all incidences were in region 3.

In Fig. 7, the fault location algorithm had a 100% of hit rate in region 1, and had a hit rate greater than 89% in the location of faults that occurred in regions 3, 7, 9 and 10. For the location of faults in region 5, a hit rate of only 58.7% was observed.

In general, it is observed that the proposed methodology had a good hit rate in the fault region location above 89.8%. It is also worth noting that, generally, when there is an error on location of faulted region, the solution presented by the algorithm is a region adjacent to the faulted region. Thus, even when a wrong fault location is made, there is a response that may assist the electrical utilities in properly identifying the place where the short-circuit occurred.
V. CONCLUSIONS

In this work, a multi-stage methodology was presented for the location of single-phase faults. The distance of the fault from the substation is initially estimated, and then the region of the fault occurrence in the system is identified. This provides a satisfactory solution to the problems of multiple estimation. A mean error rate of up to 3.97% was obtained for the estimation of fault distance. As for the identification of the region of the system in which the fault occurred, a hit rate higher than 89.8% was observed. It is worth mentioning that the performance of the algorithm varied according to the type of fault considered, since the analyzed system has single-phase lateral branches.

ACKNOWLEDGMENT

The authors would like to acknowledge the Department of Electrical and Computer Engineering, São Carlos School of Engineering, University of São Paulo (Brazil), for the research facilities provided to conduct this project. Our thanks also extend to the financial support received from CAPES and CNPq (Governmental Brazilian Institutions).

REFERENCES