Fault Locating in High Voltage Transmission Lines Based on Harmonic Components of One-End Voltage Using Random Forests

M. Farshad*, J. Sadeh*

Abstract: In this paper, an approach is proposed for accurate locating of single phase faults in transmission lines using voltage signals measured at one-end. In this method, harmonic components of the voltage signals are extracted through Discrete Fourier Transform (DFT) and are normalized by a transformation. The proposed fault locator, which is designed based on Random Forests (RF) algorithm, is trained based on these normalized harmonic components. RF algorithm has the capability of learning patterns with a large number of features. The proposed approach only requires voltage signals measured at one-end; hence, there are not problems of transmitting and synchronization of two-end data. In addition, current measurement is not required and the proposed approach is sheltered against current transformer errors and its saturation. No need for very high sampling frequency is another advantage of the proposed approach. Numerous tests carried out on a sample system indicate that accuracy of the proposed fault locator is secure against changing fault location, fault inception angle, fault resistance, and magnitude and direction of pre-fault load current. An average of 0.11% is obtained for the fault locating test errors.

Keywords: Fault location, Harmonic components, Transmission lines, Random Forests.

1 Introduction

Precise locating of permanent faults in transmission lines is crucial from the aspects of quick repairs and troubleshooting. In addition to permanent faults, temporary faults may occur in transmission lines, which are removed after re-closing of circuit breakers. Knowing exact location of temporary faults helps identifying weak points of transmission lines and to adopt appropriate measures for decreasing fault occurrence probability at those locations.

The existing approaches for fault locating in transmission lines can be classified into two main categories of traditional analytical methods and methods based on machine learning. There has been a considerable attention toward fault locating methods based on the machine learning in recent years due to the fault location problem complexity and, capability and flexibility of the learners. In the approaches based on machine learning, it is possible to train machine based on real existing patterns or patterns generated using reality-based simulation techniques. In this case, learning algorithm is responsible with task of finding hidden rules and complicated relationships between pattern features. In the fault locating approaches based on machine learning, selection and extraction of appropriate features and implementation of an efficient learning algorithm are two main issues.

Features used in fault locating may include information measured in one terminal or both terminals of transmission line. Despite the fact that using data from two ends of transmission line generally improves fault locating accuracy, need for communication channels for transmitting information of both ends and necessity for their synchronization resulted in a decline in their attraction. In fault locating approaches, it is possible to use fundamental frequency or high frequency components of voltage or/and current signals. The mentioned high frequency components will be generated due to fault occurrence in transmission lines.

There have been many machine learning tools used for fault locating in transmission lines such as Multilayer Perceptron Neural Network (MLPNN) [1-3], Radial Basis Function Neural Network (RBFNN) [4, 5], Support Vector Machine (SVM) [6-8], Extreme Learning Machine (ELM) [7], Elman Recurrent Network [9], Fuzzy Inference System (FIS) [10], Fuzzy

Iranian Journal of Electrical & Electronic Engineering, 2013.
Paper first received 31 Jan. 2013 and in revised form 6 Apr. 2013.
* The Authors are with the Department of Electrical Engineering, Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad, Khorasan-Razavi, Iran.
E-mails: m.farshad@ieee.org, sadeh@um.ac.ir.

Neural Network (FNN) [11, 12], and Adaptive Structural Neural Network (ASNN) [13]. These tools generally have desired learning and test performance for a relatively small number of input features. However, most of the mentioned learning tools are not efficiently applicable when the number of input features is large due to expansion of configuration and extreme increase in the number of learning parameters [1]. When only fundamental frequency components of voltage and current signals are implemented, the number of input features is limited. Nevertheless, when high frequency components are used, there are many features, which could be utilized. In this case, in fault locating approaches, which are designed based on the abovementioned methods, it was attempted to reduce the existing features space dimension or select a limited number of features. For example, following the applying Wavelet Transform (WT) on faulty current and voltage signals, energy and entropy of detail coefficients were calculated and then implemented as input features for dimension reduction purpose [1, 8, 9]. Moreover, Principal Component Analysis (PCA) was implemented on first and second level detail coefficients of WT of three-phase current and voltage signals and through this, input feature space dimension was reduced [13]. These methods which are implemented for dimension reduction of input feature space may result in elimination of useful features or prevention of applying crucial information in precise fault locating. For instance, application of PCA, which is an unsupervised method, is only appropriate for linear-separable data and can eliminate nonlinear relations between features [14].

In this paper, a relatively wide range of frequency components is utilized for fault locating in transmission lines. Single phase to ground fault is the most probable type, which its occurrence probability among all fault types is about 80% [15]. In this paper, a method is presented for accurate locating of single phase faults in double-ended transmission lines, which only requires voltage signals measured at one-end. In the proposed approach, amplitudes of harmonic components, which are extracted through applying DFT on voltage signals of faulted and sound phases, are used for constructing input patterns of fault locator. Due to numerous features in this issue, Random Forests (RF) algorithm is used in regression mode, which is highly efficient in dealing with problems with a large number of features [16]. RF algorithm was applied for classification of different types of faults in transmission lines, and demonstrated a desired performance [17]. But so far the performance of this algorithm has not been evaluated in the fault locating problem.

The rest of paper is organized as follows. In Section 2, a brief explanation with respect to RF algorithm is presented. In Section 3, main idea and generalities of the proposed method for single phase fault locating are presented. In Section 4, training and test patterns, which are comprised of different combinations of fault occurrence situations, are generated through simulation of a sample system. The simulations are carried out using PSCAD/EMTDC software [18]. Afterward, the proposed approach is examined using training and test patterns, and fault locating results are presented. Finally, conclusions are presented in Section 5.

2 Random Forests Algorithm

Random Forests (RF) is a learning method based on an ensemble of decision trees. Before explaining RF method, a brief description concerning learning method based on decision tree is discussed.

Each decision tree divides input space into a set of separate areas and designates a target value to each area. In basic regression mode, the target value of each area can be determined according to mean of target values of samples in that area. A greedy, top-down, recursive partitioning strategy is implemented in constructing a decision tree. In each phase of construction, a comprehensive search is carried out among all features and pertinent splitting points for obtaining maximum decline in node impurity. Simple and general algorithm of regression tree growing can be described as follows [19]:

1- Start with a node including all the existing samples (root node).
2- Explore all binary splitting of all features for the best split, which minimizes the sum of node impurities in two child nodes. Then carry out splitting and create two new child nodes based on the found best split.
3- If the maximum decrease in the node impurity is less than the predetermined threshold value δ, or the resulting child nodes contain less than q samples then stop the algorithm. Otherwise, apply step 2 to each child node.

In basic mode, in each stage of growing, only one feature and one related split point is selected and utilized. However, it is possible to take a linear combination of two or several features in consideration for splitting as well. In this case, there will be two or more features on each node [19]. In regression problems, impurity of node *nt*, $U(nt)$, is defined as follow [20]:

$$U(nt) = \frac{1}{M(nt)} \sum_{s \in nt} (y_s - \bar{y})^2 \tag{1}$$

where, $M(nt)$ indicates the number of samples located at node *nt*, y_s indicates target variable pertinent to s^{th} sample located at node *nt*, and $\bar{y}$ indicates mean of target values of all samples at node *nt*. Rate of impurity decline across a splitting on feature x_i in a node *nt* is defined as follow [20]:

$$\Delta U(x_i, nt) = U(nt) - P_L U(nt_L) - P_R U(nt_R) \tag{2}$$

where, P_L and P_R indicate a portion of samples, which proceed toward next left node nt_L and next right node nt_R.

Assigning the threshold value δ for impurity decrease and determining the minimum number of samples q in each node affect the growing of tree. If these values are taken so small, for example, if δ=0 and q=1 are assumed, then tree grows to its full potential and the number of end leaves will be equal to the number of all samples. In other words, each of the samples will be placed at an end leaf. In this case, the generalization of the tree in question will be low and overtraining will be more probable. If the setting parameters adopt relatively large values, then it is possible that tree has not grown well enough and prediction accuracy of tree is low for test samples. Generally, single decision tree is highly capable of overtraining and its prediction accuracy is commonly low as well. Instability of results obtained from single decision tree can be mentioned as another disadvantage. A small change in training patterns can result in large changes in decision tree configuration [21].

As it was mentioned previously, RF is a learning method based on an ensemble of decision trees. RF prediction model is stable while it is based on averaging results of all trees. It is worth noting that RF method is not involved with overtraining and has less generalization error [20]. In addition, it has more stability regarding the existence of noise in input data. RF method can be explained through the following brief description [20, 22, 23]:

1- A number d is specified smaller than the total number of features D.

2- For constructing each tree, a different set of existing patterns is drawn randomly with replacement. The size of the selected sample set is equal to the size of the original dataset. This approach of selection typically puts approximately one third of the existing patterns out, which are called Out-of-Bag (OOB) samples. Each tree is grown to the maximum pre-specified depth. This depth is specified based on the minimum number of samples at each end leaf. Bigger the number of the minimum samples at each end leaf selected gets, less and shorter the trees' growth and algorithm execution time, respectively.

3- At each node, a total number of d features out of D features are selected randomly.

4- For each node, the best split on the selected d features is used for creating new child nodes.

The OOB samples, which are not used in construction of specific grown tree, can be used for generalization test of the tree and also for scoring the features [23]. Averaging the number of times each feature is used at trees' splitting indicates the significance of that feature as a secondary result of trees' construction [22]. The prediction ability of RF will have more dependence on significant features and less dependence on insignificant features [23].

3 Main Concept

3.1 Harmonic Components of Voltage Signals

Fig. 1-a shows the schematic diagram of a double-ended single-circuit untransposed transmission line of 400 kV and 100 km long. Fig. 1-b also indicates the arrangement of transmission line in question, which utilized bundled conductors. Source impedance of measuring (M) and remote (R) ends in system frequency of 50 Hz are equal to Z_M=0.469+j6.283 Ω and Z_R=0.391+j5.215 Ω, respectively. The transmission line is simulated in the form of a frequency-dependent model. In simulations, sampling frequency is assumed to be 80 kHz. It is worth mentioning that the Nyquist criterion dictates using a low-pass anti-aliasing filter for the data acquisition system to avoid the frequency alias.

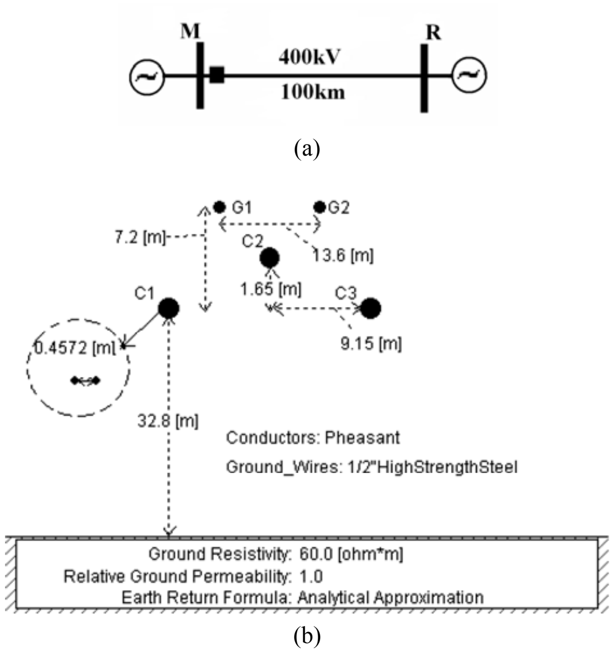

Fig. 1 Sample system under study: (a) Schematic diagram. (b) Arrangement of transmission line.

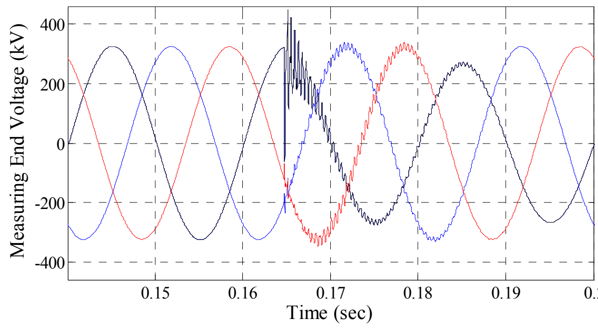

Fig. 2 Three phase voltage signals measured at the measuring end under a single phase fault.

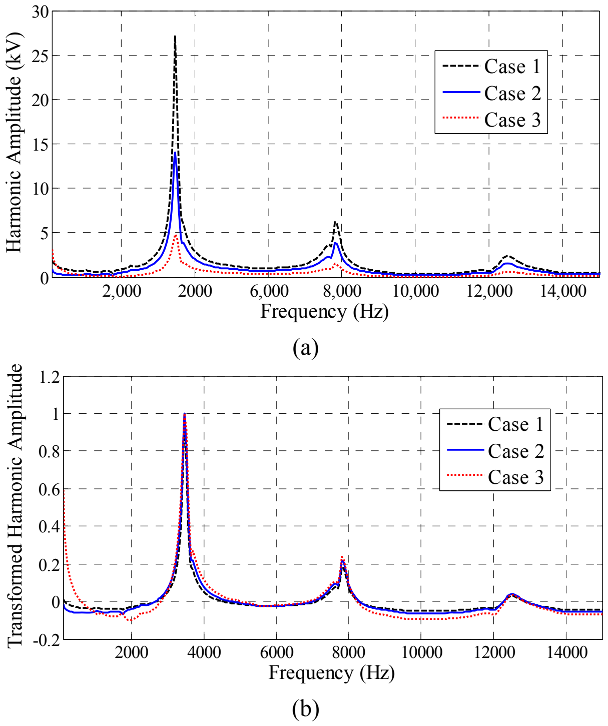

Fig. 3 Harmonic spectrum of phase 'A' voltage under A-G fault at 30 km: (a) Before transformation. (b) After transformation.

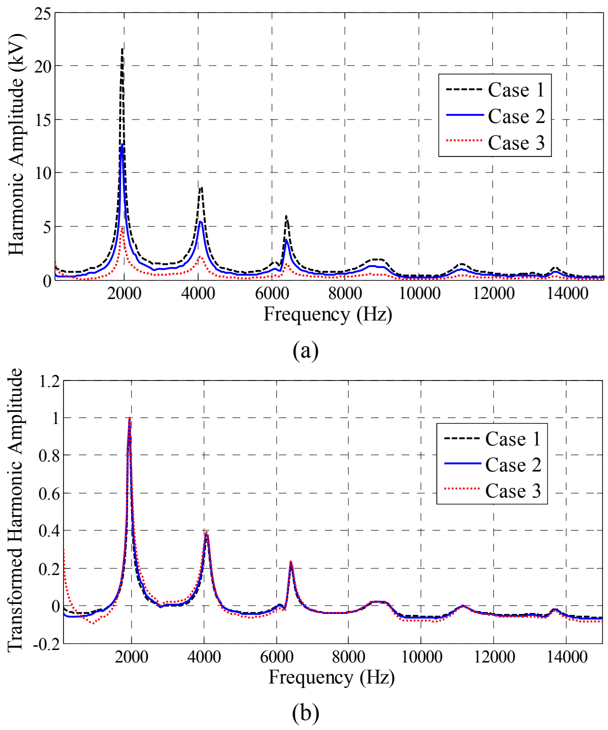

Fig. 4 Harmonic spectrum of phase 'A' voltage under A-G fault at 60 km: (a) Before transformation. (b) After transformation.

Fig. 2 demonstrates three-phase voltage signals for a single phase fault in about 0.163 s at 30 km from the measuring end. As it can be seen in this figure, there are transients in voltage signals after the fault inception. In this paper, harmonic components of voltage signals are extracted through applying DFT on one cycle of data after fault inception for fault locating purpose.

Figs. 3-a and 4-a show harmonic spectra (from 2nd to 300th order) of phase 'A' voltage under single phase to ground fault (A-G) for the following cases, at 30 km and 60 km from the measuring end, respectively:

Case 1: Fault resistance 0.01 Ω, fault inception angle 90°, and leading voltage angle of measuring end source to remote end source 20°.

Case 2: Fault resistance 5 Ω, fault inception angle 45°, and leading voltage angle of measuring end source to remote end source 30°.

Case 3: Fault resistance 10 Ω, fault inception angle 18°, and leading voltage angle of measuring end source to remote end source -30°.

In the abovementioned cases, phase 'A' voltage at the fault point is considered as the reference for fault inception angle. As it can be observed in Figs. 3-a and 4-a, in spite of changes in parameters such as fault resistance, fault inception angle and pre-fault load current, the outlines of harmonic spectra under faults at each specific distance are approximately the same. However, according to these figures, it can be found out that amplitude of each harmonic component has significant differences at the different cases. Therefore, for generating better features and normalization, the following transformation is implemented [24]:

$$V'(f_i) = \frac{V(f_i) - \frac{1}{K}\sum_{j=2}^{K} V(f_j)}{\max\limits_{f_p}\left(\left|V(f_p) - \frac{1}{K}\sum_{j=2}^{K} V(f_j)\right|\right)}, i = 2,3,\ldots,K \qquad (3)$$

where, K is the maximum order of selected harmonics, and $V(f_i)$ and $V'(f_i)$ are amplitudes of harmonic components at frequency f_i before and after transformation, respectively. The harmonic orders of the phase voltage signal from 2nd up to K^{th} are considered in the transformation. Figs. 3-b and 4-b show the transformed harmonic spectra of the faulted phase voltage signal. As it can be observed, by applying the transform on the amplitudes of harmonic components, useful features are obtained, which have the least sensitivity toward changes in the effective parameters such as fault resistance, fault inception angle, and magnitude and direction of pre-fault load current and also have high correlation with the fault location.

3.2 Single Phase to Ground Fault Locator

Single phase to ground fault locator is designed based on RF algorithm. In training phase, decision trees of RF fault locator are constructed based on training patterns. Each input pattern to RF is comprised of normalized values of harmonic component amplitudes of faulted phase and sound phases voltage signals.

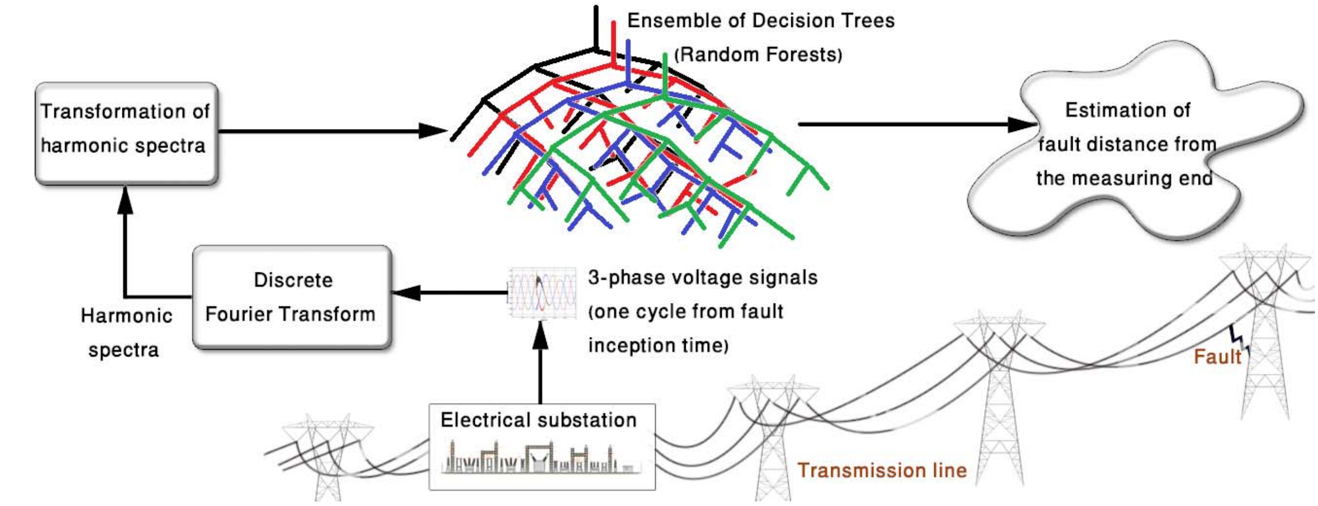

Fig. 5 Procedure of the proposed approach for single phase fault locating.

Amplitudes of harmonic components are obtained through application of DFT on one cycle of voltage signals measured at the measuring end after fault inception. After training and construction of RF trees, fault location can be acquired as output for any new patterns. The procedure of the proposed approach is presented in Fig. 5. Separate fault locators should be trained for each type of single phase faults (A-G, B-G, and C-G). It is worth noticing that the fault detection and classification are not in the scope of this paper and the time of fault signature appearance at the measuring end and the type of fault are considered as known information.

3.3 Selection of Maximum Frequency Level

For selection of appropriate maximum frequency level, harmonic spectrum of voltage under fault at a point close to measuring end should be analyzed. This issue will be examined with respect to system of Fig. 1 for better elaboration of the reason. Fig. 6 shows the transformed harmonic spectra of phase 'A' voltage under A-G fault at distances of 10 km, 50 km, and 90 km from the measuring end. As it can be observed from Fig. 6, the dominant amplitude of harmonic components related to the fault located at 10 km occurred at a higher frequency than ones of two other fault locations. Therefore, considering Fig. 6, it seems that in the system under study, the harmonic components up to frequency level of 10 kHz *(K=200)* are sufficient for generating useful features for single phase fault locator.

4 Numerical Studies

The system of Fig. 1 is modeled and simulated using PSCAD/EMTDC software [18]. Here, A-G fault is adopted out of single phase faults. MATLAB software is used for application of DFT and constructing RF.

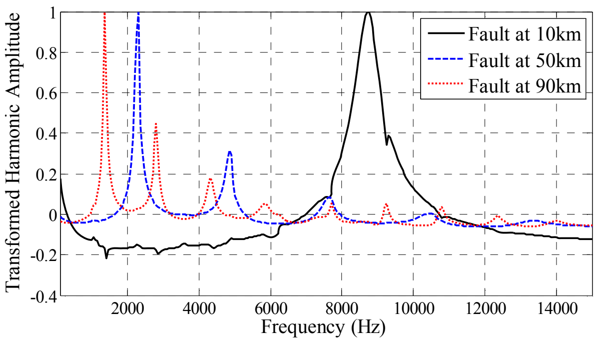

Fig. 6 Normalized harmonic spectrum of phase 'A' voltage under A-G fault at distances of 10 km, 50 km, and 90 km.

4.1 Generating Training and Test Patterns

In this stage, according to the sample system under study, training and test patterns are generated through changes of fault location, fault resistance, fault inception angle, and magnitude and direction of pre-fault load current. Generation conditions of patterns are based on a combination of various conditions of A-G fault occurrence. These conditions for training patterns are as follows:

- -Fault location varies from 10% to 90% of the line length with step of 0.5%.
- -Fault inception angle, regarding to phase 'A' voltage at the fault point as the reference, takes the values of 4.5, 9, 18, 36, 72, 108, 144, 162, 171, and 175.5 degrees.
- -Fault resistance takes the values of 0.01, 10, 30, 50, and 100 Ohms.
- -Power flow angle is taken to be 20 degrees.

The conditions which are considered for generating test patterns are as follows:

- -Faults occur at 20 different locations randomly.
- -Fault inception angle takes the values of 6.75, 13.5, 22.5, 54, 90, 126, 157.5, and 173.25 degrees.

-Fault resistance takes the values of 2, 15, 20, 40, 60, and 80 Ohms.

-Power flow angle takes the values of -30, -10, 10, and 30 degrees.

According to the analysis carried out on Fig. 6, harmonic components up to 200th order are considered for pattern generation. Therefore, regarding normalized amplitudes of three phase voltage harmonics (from 2nd to 200th) each pattern included 597 features. According to the above-mentioned conditions, a total number of 8050 training patterns ($161 \times 10 \times 5 \times 1 = 8050$), and a total number of 3840 test patterns ($20 \times 8 \times 6 \times 4 = 3840$) are generated.

4.2 Adjustment of Parameters and Training

Two parameters of RF are as follows:

-Fraction of variables which are selected randomly for each split of decision trees (*FracVar*).

-Minimum number of patterns at each end leaf of decision tree (*MinLeaf*).

For selecting desired parameters of RF, a search is conducted in discrete space of *FracVar*={10%, 20%, 40%, 60%} and *MinLeaf*={1, 3, 6} using training patterns and by growing RF up to 200 trees.

Fig. 7 illustrates the mean square error of OOB samples under different parameters of RF and growing of up to 200 trees. According to results of Fig. 7, *FracVar*=20% and *MinLeaf*=1 are selected as desired parameters for RF fault locator. Then, the A-G fault locator is formed with growing of up to 200 trees using existing training patterns and based on the selected desired parameters. It is worth mentioning that the fault locator was also formed with more than 200 trees, but in spite of an increase in the learning time, the obtained results had no significant change. Hence, it seems this selected number of trees is adequate for constructing RF fault locator in the study system. Fig. 8 demonstrates mean square error of OOB samples for the fault locator.

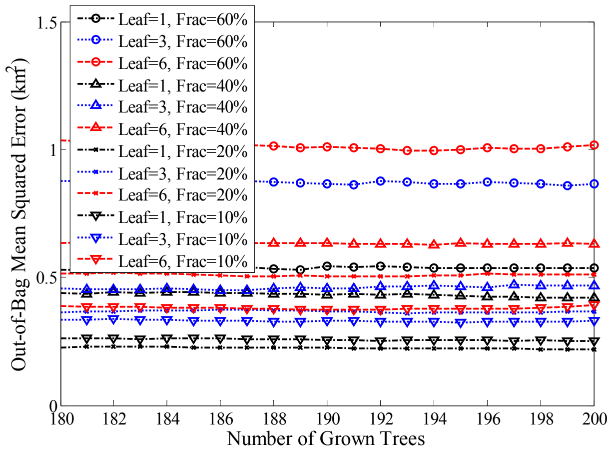

Fig. 7 Mean square error of OOB samples versus number of grown trees under different parameters.

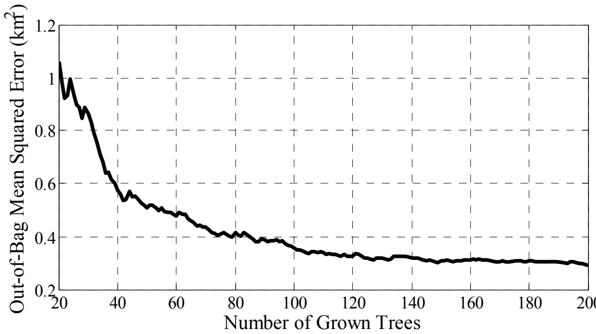

Fig. 8 Mean square error of OOB samples versus number of grown trees with selected desired parameters.

Table 1 Results of A-G fault locating for different distances of fault.

Fault location (km)	Min. error (km)	Max. error (km)	Average error (km)	Fraction of prediction errors >0.5km
10.70	0.0125	3.0275	0.2362	4.6875%
13.25	0	2.9800	0.1916	6.7708%
16.30	0.0025	5.0025	0.2203	6.7708%
21.75	0.0050	0.7350	0.1578	6.2500%
26.20	0	0.4150	0.0914	0%
31.80	0.0175	0.4500	0.1398	0%
39.30	0.0000	0.8800	0.0642	2.0833%
45.25	0.0025	0.1300	0.0432	0%
49.80	0.0025	0.4800	0.1543	0%
53.40	0.0000	0.1400	0.0388	0%
55.75	0.0000	0.3700	0.1101	0%
57.30	0.0025	0.2350	0.0723	0%
60.80	0.0100	0.4300	0.0723	0%
63.25	0	0.3075	0.1269	0%
68.75	0.0000	0.2025	0.0616	0%
71.80	0.0025	0.3775	0.1025	0%
77.20	0.0000	0.5125	0.0861	1.5625%
80.75	0.0025	0.6450	0.0359	1.0417%
87.30	0	2.2450	0.1222	2.0833%
89.25	0.0025	0.3825	0.0795	0%
All	0	5.0025	0.1103	1.5625%

4.3 Results and Discussion

The generated test patterns are presented to the trained RF. Calculation of fault location error is carried out based on difference modulus between predicted and actual values. Since length of the line under study is 100 km, the error values can be adopted in terms of percentage. The results of fault locating for different fault distances, while the fault inception angle, fault resistance, and the load current vary based on the related values in the test conditions, are shown in Table 1. In this table, there are 192 test patterns for each fault location. By examining the results of Table 1, it can be inferred that the proposed approach has sufficient accuracy regarding to the fact that generation conditions of the test patterns are different from those of the training patterns. The prediction errors have an increase for the faults occurred at the distances near to 10 km; although in these cases the average absolute errors are in an acceptable range. For investigating the under- and overestimated results of the fault locator, the raw differences between the predicted and actual fault distances are indicated in Fig. 9. As it can be seen in this figure, there are a few overestimations for the fault locations near to distance of 10 km from the measuring end, which result in an increase of the average absolute error as observed in Table 1.

Similar to Table 1, the results for each of the parameters of fault inception angle, fault resistance, and also load current are shown in Tables 2, 3, and 4, respectively. In these tables, for each specific value of fault inception angle, fault resistance, and load current, there are 480, 640, and 960 test patterns, respectively.

Table 2 Results of A-G fault locating for different fault inception angles.

Fault Inception Angle (Degree)	Min. error (km)	Max. error (km)	Average error (km)	Fraction of prediction errors >0.5km
6.75	0	2.2850	0.1234	2.2917%
13.5	0	0.2450	0.0707	0%
22.5	0	0.2675	0.0779	0%
54	0	0.2700	0.0802	0%
90	0	0.3125	0.0802	0%
126	0.0025	0.2975	0.0798	0%
157.5	0	0.3225	0.0777	0%
173.25	0	5.0025	0.2928	10.2083%

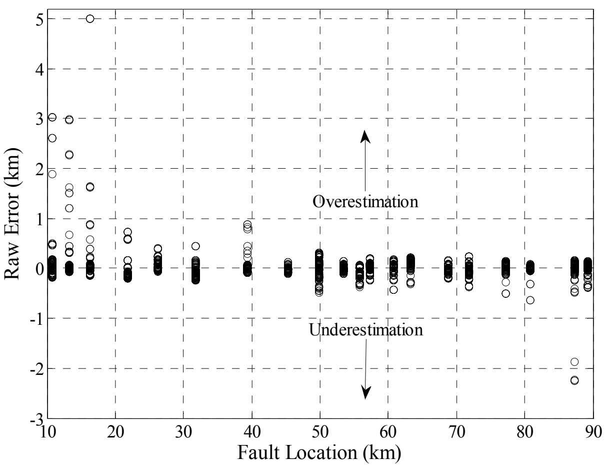

Fig. 9 Raw differences between the predicted and actual fault distances for different fault locations.

Table 3 Results of A-G fault locating for different fault resistances.

Fault Resistance (Ω)	Min. error (km)	Max. error (km)	Average error (km)	Fraction of prediction errors >0.5km
2	0	2.2450	0.0970	1.0938%
15	0	0.6450	0.0721	0.3125%
20	0	0.3275	0.0685	0%
40	0	0.7350	0.0810	0.6250%
60	0.0025	2.6175	0.1535	3.9063%
80	0	5.0025	0.1900	3.4375%

Table 4 Results of A-G fault locating for different load currents.

Power Flow Angle (Degree)	Min. error (km)	Max. error (km)	Average error (km)	Fraction of prediction errors >0.5km
-30	0	5.0025	0.1107	1.5625%
-10	0	5.0025	0.1102	1.6667%
10	0	5.0025	0.1102	1.6667%
30	0	5.0025	0.1103	1.3542%

Considering Table 2, it can be concluded that accuracy of the proposed fault location algorithm at fault inception angles close to zero has a slight decline. However, it shows a remarkable accuracy at other angles. At worst case scenario, at angle 173.25^{o}, the average absolute error value equals 292.8 m. If single phase fault occurs when the voltage of faulted phase at fault location is near to zero, then the voltage signals measured at the terminal will not be rich in terms of harmonic contents. Under such circumstances, the proposed approach may not have a desirable performance. This is a common trouble spot for fault locating methods which are based on fault generated high frequency transients [25]. In the system under study, if fault inception angle distance from zero crossing point of fault location voltage is less than 4.5^{o}, then the proposed approach may not show an acceptable performance. If these areas are taken in 360^{o} of a cycle, their lengths add up to 18^{o}. As a result, fault occurrence probability in these areas is 5%. In the other words, the proposed approach has an acceptable performance at 95% of the time.

Through examination of Table 3, it can be observed that by increasing the value of fault resistance, prediction accuracy has a slight decrease. At the worst case scenario, in which the fault resistance is 80 Ω, the average absolute value of errors is 190 m.

Based on the results of Table 4, it can be inferred that the proposed approach roughly has no reliance on magnitude and direction of pre-fault current.

4.4 Distinctive Aspects

It is shown that using the normalized amplitudes of harmonic components of three phase voltage signals is highly efficient for single phase fault locating. However, in this case, number of used features is considerable (597 features for the system under study). Common tools and methods used for fault locating such as Artificial Neural Networks and Support Vector Machines are not practically capable of appropriate learning of patterns with a large number of features. On the other hand, dimension reduction using linear transforms or experimental selection of several features may result in elimination or disregard of some informative features. Consequently, in this paper, RF learning algorithm is utilized in regression mode, which has appropriate computational efficiency and shows a good performance dealing with patterns with numerous features.

Using only voltage signals of one-end and no need for very high sampling frequency are the other advantages of the proposed method. In the proposed method, problems caused by saturation of current transformers are non-existent. In a recently published article [26], a method was presented for ground fault locating in transmission lines, which was based on high frequency transients of voltage signals measured at one-end. This method needs a high sampling frequency and its locating accuracy will decrease with reducing the sampling frequency. As a confirmation of this claim, the presented results for locating the single phase to ground fault at the middle of a transmission line of 230 kV and 100 km in length, in sampling frequencies of 6700 kHz and 670 kHz showed the average estimation errors of 360 m and 6700 m, respectively [26].

5 Conclusion

In this paper, an approach is proposed for single phase fault locating using harmonic components of one-end voltage signals. In the suggested approach, Random Forests (RF) learning algorithm is used, which demonstrates an appropriate capability with respect to a large number of input features. The results obtained from numerical studies are indicative of the efficiency of the proposed approach addressing problems of single phase fault locating issue. The obtained average absolute error value for test patterns with simultaneous changes in fault resistance, fault inception angle, and magnitude and direction of load current is low as 0.11%. Some positive points of the selected features can be mentioned as their low sensitivity to changes of pre-fault load current, fault resistance, and fault inception angle, and also high dependence of their values on the fault location. In addition, lack of problems caused by transmitting and synchronizing of two-end data, lack of problems caused by saturation of current transformers during fault occurrence, and also no need for a very high sampling frequency can be pointed out as the other salient advantages of the proposed approach.

References

[1] Ekici S., Yildirim S. and Poyraz M., “Energy and entropy-based feature extraction for locating fault on transmission lines by using neural network and wavelet packet decomposition”, *Expert Systems with Applications*, Vol. 34, No. 4, pp. 2937-2944, May 2008.

[2] Bhowmik P. S., Purkait P. and Bhattacharya K., “A novel wavelet transform aided neural network based transmission line fault analysis method”, *Electrical Power and Energy Systems*, Vol. 31, No. 5, pp. 213-219, Jun. 2009.

[3] Ezquerra J., Valverde V., Mazon A. J., Zamora I. and Zamora J. J., “Field programmable gate array implementation of a fault location system in transmission lines based on artificial neural networks”, *IET Generation, Transmission & Distribution*, Vol. 5, No. 2, pp. 191-198, Feb. 2011.

[4] Joorabian M., Taleghani-Asl S. M. A. and Aggarwal R. K., “Accurate fault locator for EHV transmission lines based on radial basis function neural networks”, *Electric Power Systems Research*, Vol. 71, No. 3, pp. 195-202, Nov. 2004.

[5] Samantaray S. R., Dash P. K. and Panda G., “Fault classification and location using HS-transform and radial basis function neural network”, *Electric Power Systems Research*, Vol. 76, No. 9-10, pp. 897-905, Jun. 2006.

[6] Salat R. and Osowski S., “Accurate fault location in the power transmission line using support vector machine approach”, *IEEE Transactions on Power Systems*, Vol. 19, No. 2, pp. 979-986, May 2004.

[7] Malathi V., Marimuthu N. S. and Baskar S., “Intelligent approaches using support vector machine and extreme learning machine for transmission line protection”, *Neurocomputing*, Vol. 73, No. 10-12, pp. 2160-2167, Jun. 2010.

[8] Ekici S., “Support vector machines for classification and locating faults on transmission lines”, *Applied Soft Computing*, Vol. 12, No. 6, pp. 1650-1658, Jun. 2012.

[9] Ekici S., Yildirim S. and Poyraz M., “A transmission line fault locator based on Elman recurrent networks”, *Applied Soft Computing*, Vol. 9, No. 1, pp. 341-347, Jan. 2009.

[10] Reddy M. J. and Mohanta D. K., “A wavelet-fuzzy combined approach for classification and location of transmission line faults”, *Electrical Power and Energy Systems*, Vol. 29, No. 9, pp. 669-678, Nov. 2007.

[11] Jung C. K., Kim K. H., Lee J. B. and Klockl B., “Wavelet and neuro-fuzzy based fault location for combined transmission systems”, *Electrical Power and Energy Systems*, Vol. 29, No. 6, pp. 445-454, Jul. 2007.

[12] Sadeh J. and Afradi H., “A new and accurate fault location algorithm for combined transmission lines using Adaptive Network-Based Fuzzy Inference System”, *Electric Power Systems Research*, Vol. 79, No. 11, pp. 1538-1545, Nov. 2009.
[13] Jiang J. A., Chuang C. L., Wang Y. C., Hung C. H., Wang J. Y., Lee C. H. and Hsiao Y. T., “A hybrid framework for fault detection, classification, and location-part I: concept, structure, and methodology”, *IEEE Transactions on Power Delivery*, Vol. 26, No. 3, pp. 1988-1998, Jul. 2011.
[14] Mora-Florez J., Morales-Espana G. and Perez-Londono S., “Learning-based strategy for reducing the multiple estimation problem of fault zone location in radial power systems”, *IET Generation, Transmission & Distribution*, Vol. 3, No. 4, pp. 346-356, Apr. 2009.
[15] Beiza J., Hosseinian S. H. and Vahidi B., “Fault type estimation in power systems”, *Iranian Journal of Electrical & Electronic Engineering*, Vol. 5, No. 3, pp. 185-195, Sep. 2009.
[16] Kitchen C. M. R., “Nonparametric variable selection using machine learning algorithms in high dimensional (large P, small N) biomedical applications”, in: Laskovski, A. N. (Ed.), *Biomedical Engineering: Trends in Electronics, Communications and Software*, Rijeka, Croatia: InTech, 2011, pp. 589-600.
[17] Jamehbozorg A. and Shahrtash S. M., “A decision-tree-based method for fault classification in single-circuit transmission lines”, *IEEE Transactions on Power Delivery*, Vol. 25, No. 4, pp. 2190-2196, Oct. 2010.
[18] Power Systems Computer Aided Design/ Electromagnetic Transients including Direct Current (PSCAD/EMTDC), Computer Software, *Manitoba HVDC Research Centre*, Winnipeg (Canada), 2005.
[19] Loh W. Y., “Classification and regression trees”, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol. 1, No. 1, pp. 14-23, Jan./Feb. 2011.
[20] Tuv E., Borisov A. and Torkkola K., “Ensemble-based variable selection using independent probes”, in: Liu, H. and Motoda, H. (Eds.), *Data mining and knowledge discovery-computational methods of feature selection*, Boca Raton, Florida: Chapman & Hall/CRC, Taylor & Francis Group, 2008, pp. 131-145.
[21] Torgo L. F. R. A., “Inductive Learning of Tree-based Regression Models”, PhD Thesis, Department of Computer Science, Faculty of Sciences, University of Porto, 1999.
[22] Tuv E., “Ensemble learning”, in: Guyon, I., Gunn, S., Nikravesh, M. and Zadeh, L. A. (Eds.), *Studies in fuzziness and soft computing 207-feature extraction: foundations and applications*, Heidelberg, Berlin: Springer-Verlag, 2006, pp. 187-204.
[23] Breiman L., “Random forests”, *Machine Learning*, Vol. 45, pp. 5-32, 2001.
[24] Hestilow T. J. and Huang Y., “Clustering of gene expression data based on shape similarity”, *EURASIP Journal on Bioinformatics and Systems Biology*, pp. 1-12, 2009.
[25] Jafarian P. and Sanaye-Pasand M., “A traveling-wave-based protection technique using wavelet/PCA analysis”, *IEEE Transactions on Power Delivery*, Vol. 25, No. 2, pp. 588-599, Apr. 2010.
[26] Mardiana R., Al-Motairy H. and Su C. Q., “Ground fault location on a transmission line using high-frequency transient voltages”, *IEEE Transactions on Power Delivery*, Vol. 26, No. 2, pp. 1298–1299, Apr. 2011.

Mohammad Farshad was born in Gonbad-e-Qabus, Iran, in 1981. He received the B.Sc. degree in power transmission and distribution networks engineering from Power and Water University of Technology (PWUT), Tehran, Iran, in 2003, and the M.Sc. degree in power system engineering from Ferdowsi University of Mashhad, Mashhad, Iran, in 2006, where he is currently pursuing the Ph.D. degree in power system engineering. His main research interests are power system protection and operation.

Javad Sadeh was born in Mashhad, Iran, in 1968. He received the B.Sc. and M.Sc. degrees in electrical engineering (Hons.) from Ferdowsi University of Mashhad, Mashhad, Iran, in 1990 and 1994, respectively, and the Ph.D. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, with the collaboration of the electrical engineering laboratory of the Institut National Polytechnique de Grenoble (INPG), Grenoble, France, in 2001. Currently, he is an Associate Professor in the Department of Electrical Engineering, Ferdowsi University of Mashhad. His research interests are power system protection, dynamics, and operation.