A Novel Fault-Location Method for HVDC Transmission Lines Based on Similarity Measure of Voltage Signals

Mohammad Farshad, *Student Member, IEEE*, and Javad Sadeh

***Abstract*—In this paper, a method for fault locating in HVDC transmission lines is proposed which only uses the voltage signal measured at one of the line terminals. The postfault voltage signal, in a relatively short-time window, is considered and the corresponding fault location is estimated based on the similarity of the captured voltage signal to existing patterns. In this approach, the Pearson correlation coefficient is used to measure the similarity. Despite simplicity and low complexity of the proposed fault-location method, it does not suffer from the technical problems which are associated with the traveling-wave-based methods, such as the difficulty of identifying traveling wavefronts or the strong dependency of accuracy on the sampling frequency. Numerous training and test patterns are obtained by simulating various fault types in a long overhead HVDC transmission line under different fault location, fault resistance, and prefault current values. The accuracy of the proposed fault-location method is verified using these patterns.**

***Index Terms*—HVDC system, pattern recognition, similarity measure, single-end fault-location method.**

I. Introduction

HVDC systems provide an opportunity to transfer large amounts of power over long distances more economically and with lower losses compared to HVAC transmission systems. Accurately locating permanent and nonpermanent faults in long HVDC transmission lines is very important in corrective and preventive maintenance operations speedup and maintaining continuity of power transmission. Most of the methods, which have been proposed so far for fault locating in HVDC transmission lines, are based on the traveling-wave theory [1]–[7]. The traveling-wave-based methods are very accurate and valuable studies have been performed to improve their performance considering different system configurations, such as the combination of overhead lines and cables [1], and the star-connected multiterminal topologies [2]. However, as mentioned in [8], the traveling-wave-based methods suffer from inherent problems, some of which include the necessity of having experience and skill for wavefront identification, and the high dependency of accuracy on the sampling rate. In this regard, a nontraveling-wave time-domain fault-location principle has been proposed in [8]. In the method of [8], the voltage profile along the line was calculated using two-terminal voltage and current measurements and the corresponding fault location was identified from the obtained voltage distribution.

As an alternative to the existing methods, a fault-location method based on the machine-learning and the pattern-recognition techniques can demonstrate appropriate performance and flexibility in different conditions when encountering inherent characteristics and specifications of HVDC systems. When using strategies based on the machine-learning and the pattern-recognition techniques, the key point is to determine appropriate input features and choose an algorithm and structure in accordance with the specific problem to be solved.

Single-ended or double-ended measurements can be used for fault locating in HVDC transmission lines. Usually, the methods based on double-ended measurements [1]–[4], [8] are more accurate. However, more reliability in access to required measured data, no need for transmitting and synchronizing measurements of both ends, less complexity, and lower cost are some of the advantages of the methods which are based on single-ended measurements [4]–[7]. Voltage and current signals are the commonly available records in the line terminals. Founding the fault-location method based on just one of these signals can be effective in terms of preventing the combination of measurement errors and increasing the reliability in access to required data.

In this paper, a learning-based method for fault locating in HVDC transmission lines is proposed. In the proposed method, samples from a time window of the postfault voltage signal measured at one of the line terminals are considered as the input features. The fault location is estimated by measuring the similarity between the input pattern and existing patterns based on the Pearson correlation coefficient.

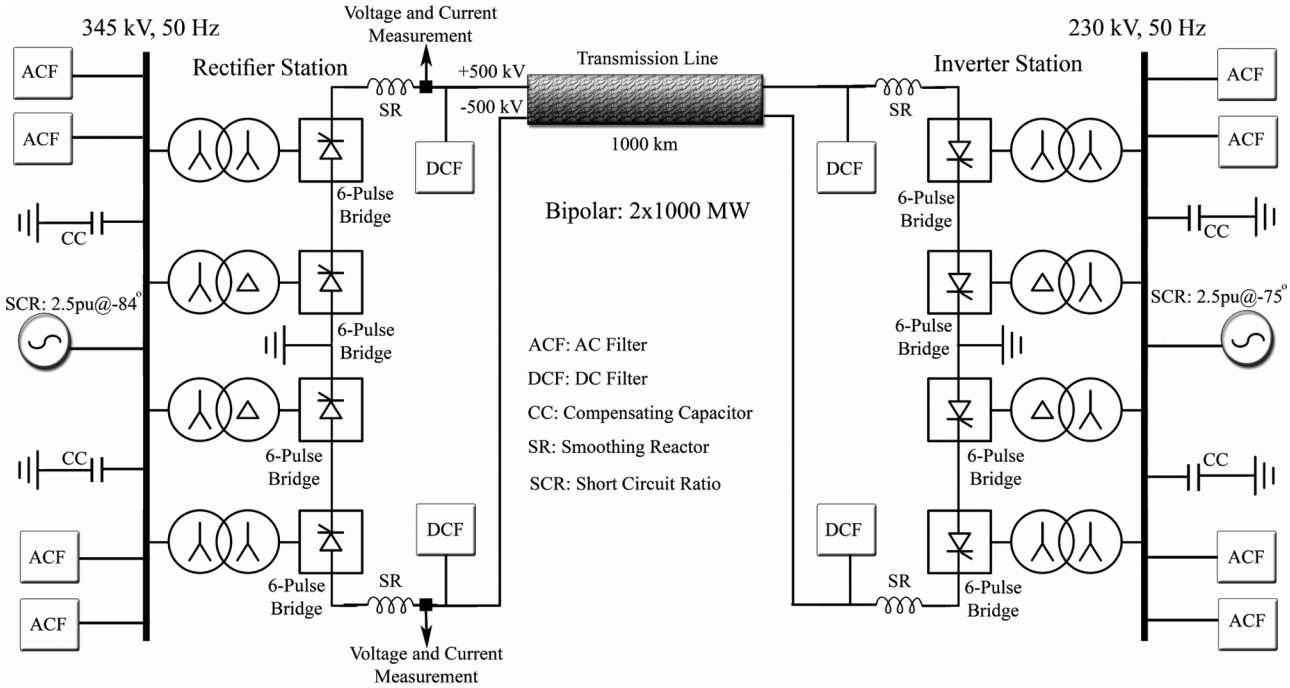
The rest of this paper is organized as follows. In Section II, the main idea and concept of the proposed fault-location approach are explained. In Section III, numerous training and test patterns are generated through simulating different fault types in a sample bipolar HVDC system with different values for the fault location, fault resistance, and the prefault current. Then, the proposed method is applied to the training and test patterns, and the fault-location results are presented. Finally, concluding remarks are given in Section IV.

Manuscript received February 08, 2013; revised May 20, 2013; accepted July 02, 2013. Date of publication July 18, 2013; date of current version September 19, 2013. Paper no. TPWRD-00176-2013.

The authors are with the Electrical Engineering Department, Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad 91779-48944, Iran (e-mail: m.farshad@ieee.org; sadeh@um.ac.ir).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPWRD.2013.2272436



Note: The SCR values are calculated as the ratio between the AC system short circuit power and the rated DC power (2000 MW).

Fig. 1. Single-line diagram of the sample system under study, adopted from the CIGRE benchmark [10].

II. Main Idea and Concept

A. Voltage Signals and Shape Similarity

Fig. 1 illustrates the single-line diagram of an HVDC system with the nominal voltage of ±500 kV and the rated transmission power of 2000 MW. The parameters of this bipolar system, which are simulated through PSCAD/EMTDC software, [9] are adopted from the CIGRE monopolar system [10], except for the transmission line. It is worth noticing that the 1000-km-long overhead transmission line of the system is simulated using the frequency-dependent model and based on the configuration of Fig. 2(a)[11]. Also, the equivalent circuit of Fig. 2(b) is used to model the various fault types. In this figure, each $R_{\rm f}$ indicates two-state switching resistance which can be "on" or "off" to model different fault types. Each $R_{\rm f}$ has a very large resistance and is nearly open circuit in the "off" state; and in the "on" state, it has a resistance equal to the specified fault resistance value. It worth noting that the sampling frequency of signals in the system under study is adopted to be 80 kHz.

Fig. 3 shows the dc voltage and current signals measured at the rectifier station for a positive-pole-to-ground (PG) fault occurring in about 0.45 s and at 100 km from the measuring point. It should be noted that there are two main factors, including the oscillations due to capacitor energy storage [12] and the fault-generated transients due to reflections of traveling waves [13], that can affect the shape of the postfault voltage signal while having a sensible correlation with the fault distance.

Fig. 4 shows the positive-pole voltage signals after PG faults at the distance of 100, 250, and 400 km from the rectifier station (measuring point) in the following cases:

Case 1) fault resistance of 0.01 Ω and prefault current of 1200 A in the HVDC line.

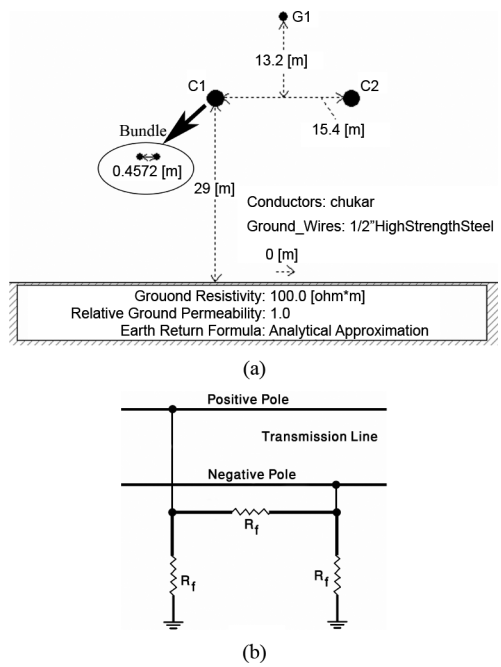


Fig. 2. (a) The configuration of transmission line. (b) The equivalent circuit in modeling different fault types.

Case 2) fault resistance of 30 Ω and prefault current of 1200 A in the HVDC line.

Case 3) fault resistance of 10 Ω and prefault current of 600 A in the HVDC line.

Case 4) fault resistance of 50 Ω and prefault current of 1800 A in the HVDC line.

The postfault signals illustrated in Fig. 4 are captured from the moment of voltage drop to less than 400 kV. As can be seen

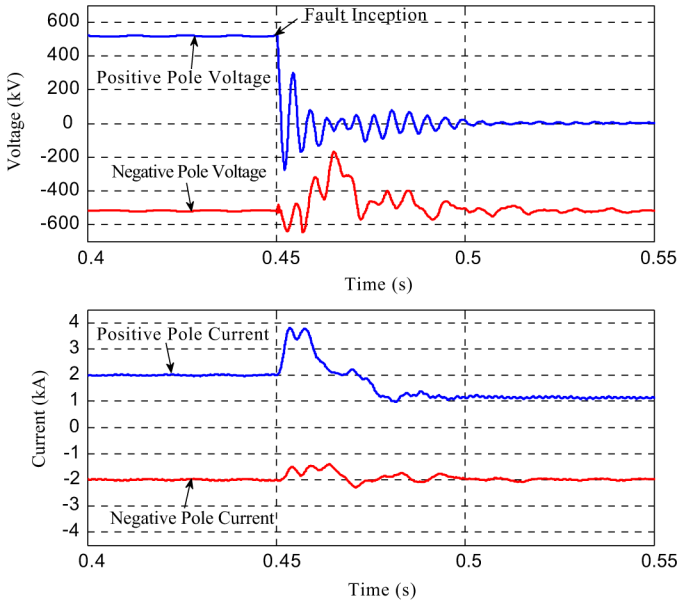


Fig. 3. DC voltage and current signals for a PG fault at 100 km from the rectifier station (measuring point).

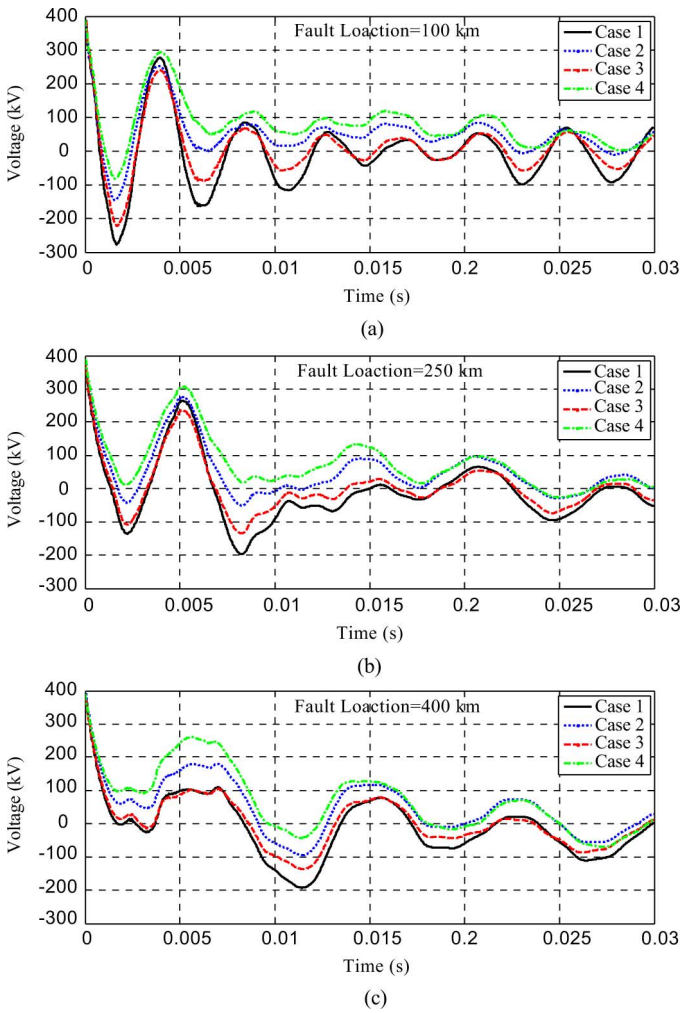


Fig. 4. Positive-pole voltage signals after PG faults at the distance of: (a) 100, (b) 250, and (c) 400 km.

from this figure, the postfault voltage signals of different fault distances have dissimilar shapes. Also, with little attention, it can be observed that despite the different conditions of Case 1 to Case 4, the shapes of voltage signals are almost similar for each specific fault location. In summary, it can be comprehended that despite the variations in fault resistance and prefault current, the overall shape of the voltage signal is almost the same for a specific fault location; but it changes with fault location. This characteristic can be used for fault locating. To this end, a criterion should be defined for measuring the shape similarity of signals. The Pearson correlation coefficient is one of the efficient similarity measures which can be defined as follows for the two signal samples $x = \{x_1, x_2, \ldots, x_n\}$ and $y = \{y_1, y_2, \ldots, y_n\}$[14]

$$r(x,y) = \frac{\sum_{i=1}^{n}(x_i - \frac{1}{n}\sum_{j=1}^{n}x_j)(y_i - \frac{1}{n}\sum_{j=1}^{n}y_j)}{\sqrt{\sum_{i=1}^{n}(x_i - \frac{1}{n}\sum_{j=1}^{n}x_j)^2}\sqrt{\sum_{i=1}^{n}(y_i - \frac{1}{n}\sum_{j=1}^{n}y_j)^2}} \quad (1)$$

where $r(x,y)$ is the Pearson correlation coefficient of the two signals x and y, and $r(x,y) \in [-1,+1]$. The value $+1$ for r indicates perfect positive linear correlation; the value 0 indicates no linear correlation at all, and the value -1 indicates perfect negative linear correlation. In fact, a larger value of r for two signals indicates more correlation and similarity between them. Table I presents the r values for the signal pairs illustrated in Fig. 4. The results presented in this table confirm the intuitive results obtained from the visual inspection of the voltage waveforms.

B. Proposed Fault-Location Algorithm

The voltage signals required for fault locating are measured at one terminal station of HVDC systems, in front of the smoothing reactor at the dc line side. There are published works focusing on conventional [15] and valuable new methods [16], [17] for fault detection and classification in HVDC systems which can be utilized before executing the proposed fault-location algorithm. In the proposed method, after fault detection and classification, which are not in the scope of this paper, the voltage signal of the faulted pole, which has been recorded in the buffer memory, is used to extract the required samples. It is worth noting that in the case of the faults that involve both of the line poles, one of the faulted poles, for example the positive pole, is selected and its voltage signal is used for fault locating. In the proposed fault-location approach, required voltage samples are gathered from the moment that the absolute value of voltage drops below the threshold value V_{thd} until 10 ms later. In fact, these samples are the pattern's features. After extracting the required signal samples and generating the new pattern, the similarity of this pattern to each of existing patterns in the database is measured, and the corresponding fault location is estimated by weighted average of the target values corresponding to the k most similar patterns. It is clear that the proposed fault-location method is founded on the k-nearest neighbor (k-NN) principle. Therefore, the similarity measure should be converted to a positive valued distance metric. The following equation is used for this purpose [18]:

$$d_r(x,y) = 1 - r(x,y) \quad (2)$$

TABLE I
CORRELATION COEFFICIENTS FOR THE SIGNAL PAIRS OF FIG. 4

		Case 1			Case 2		
	Fault location	100 km	250 km	400 km	100 km	250 km	400 km
Case 3	100 km	0.9816	0.3418	0.2820	0.9705	0.3967	0.2667
	250 km	0.2661	0.9904	0.5751	0.4460	0.9775	0.5566
	400 km	0.2232	0.5346	0.9901	0.3315	0.6736	0.9817
Case 4	100 km	0.8340	0.5008	0.4146	0.9559	0.5956	0.4438
	250 km	0.2773	0.8493	0.6991	0.4884	0.9545	0.7477
	400 km	0.1627	0.4727	0.8854	0.3334	0.6532	0.9530

where $d_{\mathrm{r}}(x,y)$ is the Pearson distance of the signals x and y, and $d_r(x,y) \in [0,2]$. Greater values of the similarity measure r, that is , closer to $+1$, lead to smaller values of the distance d_r. The fault location corresponding to the new pattern is estimated based on the following equation [19]:

$$\widehat{F}_L(x') = \frac{\sum\limits_{x \in S(x')} F_L(x).e^{-(d_r(x,x'))^2}}{\sum\limits_{x \in S(x')} e^{-(d_r(x,x'))^2}} \tag{3}$$

where $F_{\mathrm{L}}(x)$ is the corresponding fault location of the existing pattern x, $\widehat{F}_L(x')$ is the estimated fault location corresponding to the new pattern x', $S(x')$ is the set of k patterns most similar to the new pattern x', and $d_r(x,x')$ is the Pearson distance of the existing pattern x and the new pattern x'.

It is worth noting that in the proposed method, a separate database of training patterns should be provided for each fault type and they shall be used based on the type of occurring fault.

In summary, the proposed fault-location algorithm consists of the following steps:

Step 1) Receiving the fault-type information as well as the buffered voltage signal of the faulted pole.

Step 2) Determining the moment in which the absolute value of voltage sample drops below the threshold value V_{thd} and gathering the voltage samples from this moment until 10 ms later.

Step 3) Generating the new pattern using the extracted voltage samples as the features.

Step 4) Determining the k most similar patterns to the new pattern, by similarity measuring and searching through the database corresponding to the occurred fault type.

Step 5) Estimating the fault location corresponding to the new pattern based on the weighted average of corresponding fault locations of the k most similar patterns [using (3)].

III. NUMERICAL STUDIES

Here, the location of the positive-pole-to-ground (PG), positive-pole-to-negative-pole (PN), and positive-pole-to-negative-pole-to-ground (PNG) faults in the system of Fig. 1 are considered. The PSCAD/EMTDC software [9] is utilized for the system simulation, and the MATLAB environment [20] is used for implementation of the proposed fault-location algorithm.

A. Generating the Training and Test Patterns

In this phase, training and test patterns are generated through simulating various fault types in the sample system, and by changing the fault location, fault resistance, and prefault current. In fact, these patterns are generated based on combinations of various conditions of the PG, PN, and PNG faults; theses conditions for training cases are as follows.

- Fault distance from the measuring end varies from 10 to 990 km with the step of 2 km.
- Fault resistance takes the values of 0.01, 10, 30, 50, and 100 Ω.
- Prefault current of the HVDC line takes the values of 600, 1200, and 1800 A.

Also, the following conditions are considered in generation of test patterns:

- Faults occur at 25 different locations randomly.
- Fault resistance takes the values of 2, 15, 20, 40, 60, and 80 Ω.
- Prefault current of the HVDC line takes the values of 800, 1000, 1400, and 1600 A.

All of the test and training patterns are generated using the method provided in Steps 1)–3) of the proposed algorithm (see Section II-B). Since all fault types considered in this study involve the positive pole, only the positive-pole voltage measured at the rectifier station is used in generating the training and test patterns. Also, the threshold value V_{thd}, which is used in determining the beginning moment of the 10-ms data window, is set to 400 kV (80% of the nominal dc voltage). Considering the sampling frequency of 80 kHz and the 10-ms data window, it can be comprehended that each generated pattern consists of 800 features.

B. Fault-Location Results

Prior to testing the fault locator, the value of k is determined through applying the 10-fold cross-validation process on the training patterns. After applying this process on the training patterns, the obtained k value for all three fault types was 4. After adjusting the k value for each fault type, Steps 4) and 5) of the proposed algorithm (see Section II-B) are performed on the test patterns, and the fault-location results are provided.

The percentage error in the fault-location estimate e is computed by using the following equation and considered as the performance index:

$$e = \frac{|L_{\mathrm{est}} - L_{\mathrm{act}}|}{L_T} \times 100 \tag{4}$$

TABLE II
FAULT-LOCATION RESULTS FOR THE DIFFERENT FAULT TYPES AND FAULT DISTANCES

	Percentage fault-location errors- PG faults				Percentage fault-location errors- PN faults				Percentage fault-location errors- PNG faults			
Fault distance (km)	Min. error (%)	Max. error (%)	Average error (%)	Fraction of errors >1%	Min. error (%)	Max. error (%)	Average error (%)	Fraction of errors >1%	Min. error (%)	Max. error (%)	Average error (%)	Fraction of errors >1%
11.05	0.0048	0.0949	0.0439	0%	0.0046	0.0552	0.0319	0%	0.0046	0.0950	0.0488	0%
55.10	0.0099	0.2600	0.0704	0%	0.0100	0.0900	0.0313	0%	0.0099	0.1100	0.0396	0%
99.15	0.0149	0.1850	0.0558	0%	0.0150	0.0850	0.0421	0%	0.0150	0.1150	0.0483	0%
117.20	0.0200	0.2799	0.0758	0%	0.0200	0.0800	0.0442	0%	0.0200	0.0800	0.0383	0%
147.05	0.0049	0.4050	0.1467	0%	0.0049	0.1950	0.0487	0%	0.0049	0.1550	0.0383	0%
192.95	0.0050	0.4450	0.0963	0%	0.0049	0.4050	0.0546	0%	0.0049	0.0950	0.0308	0%
231.00	0.0000	1.4000	0.2333	4.1667%	0.0000	0.2000	0.0375	0%	0.0000	0.2000	0.0562	0%
268.90	0.0100	1.5900	0.3225	4.1667%	0.0100	0.1100	0.0313	0%	0.0100	0.1100	0.0312	0%
281.05	0.0050	1.6050	0.3350	8.3333%	0.0050	0.1450	0.0367	0%	0.0050	0.0950	0.0225	0%
303.10	0.0100	0.7600	0.2542	0%	0.0100	0.1900	0.0392	0%	0.0100	0.0600	0.0221	0%
347.30	0.0300	1.1700	0.3833	4.1667%	0.0200	0.1699	0.0442	0%	0.0200	0.0700	0.0308	0%
381.05	0.0050	0.8050	0.2796	0%	0.0050	0.1949	0.0337	0%	0.0050	0.0950	0.0137	0%
409.00	0.0000	1.3500	0.3542	8.3333%	0.0000	0.1000	0.0354	0%	0.0000	0.1000	0.0208	0%
461.50	0.0500	0.8000	0.3583	0%	0.0000	0.1500	0.0750	0%	0.0000	0.2000	0.0521	0%
490.40	0.0100	1.3900	0.3121	4.1667%	0.0100	0.2599	0.0879	0%	0.0100	0.2600	0.0575	0%
591.35	0.0650	1.1150	0.5267	4.1667%	0.0150	0.2350	0.0867	0%	0.0150	0.2350	0.0917	0%
631.00	0.0000	0.9500	0.4812	0%	0.0000	0.2000	0.0646	0%	0.0000	0.3999	0.0708	0%
667.15	0.0150	1.2150	0.4604	4.1667%	0.0149	0.2150	0.0779	0%	0.0149	0.2150	0.0767	0%
692.95	0.1950	0.8450	0.5525	0%	0.0050	0.2049	0.0550	0%	0.0050	0.2049	0.0708	0%
729.30	0.1200	0.8700	0.4283	0%	0.0300	1.0300	0.2871	4.1667%	0.0300	0.5700	0.2600	0%
777.70	0.0200	0.8800	0.2975	0%	0.0300	0.7701	0.2121	0%	0.0300	0.6700	0.2392	0%
799.00	0.0000	1.0500	0.3083	4.1667%	0.0000	0.5500	0.1438	0%	0.0000	0.5500	0.1708	0%
801.25	0.0250	1.1750	0.3375	4.1667%	0.0250	0.4250	0.1979	0%	0.0250	0.4250	0.1604	0%
951.00	0.0000	1.1000	0.4167	4.1667%	0.0000	0.2000	0.1021	0%	0.0000	0.2000	0.0896	0%
987.05	0.0049	0.2550	0.0729	0%	0.0049	0.3050	0.0429	0%	0.0049	0.2050	0.0350	0%
All	0.0000	1.6050	0.2881	2.1667%	0.0000	1.0300	0.0777	0.1667%	0.0000	0.6700	0.0726	0%

where $L_{\rm est}$ is the estimated fault distance from the measuring end, $L_{\rm act}$ is the actual fault distance from the measuring end, and L_T is the total length of the transmission line.

Based on the conditions mentioned in Section III-A, regarding each fault type, there are a total number of 24 test patterns for each fault distance. Table II provides the estimation results for different fault distances from the rectifier station, including the minimum, maximum, and average of percentage errors and the fraction of percentage errors bigger than 1%. As can be seen from this table, the overall average errors for the PG, PN, and PNG faults are 0.2881%, 0.0777%, and 0.0726%, respectively. Also, it can be observed that the estimation errors of the proposed fault-location method are in a fairly acceptable range, and the overall fractions of fault-location errors bigger than 1% are as low as 2.1667%, 0.1667%, and 0% for the PG, PN, and PNG faults, respectively. The average errors for the PN and PNG faults are far less than those for the PG faults; it means that in the PN and PNG fault cases, the correlation of voltage signals in each fault location, despite the changes in the fault conditions, is more than that of the PG fault cases. It is worth noting that in the PN and PNG fault cases, the fault voltage level is twice that in the PG fault cases. If the training set is extended by considering more fault locations, fault resistances, and prefault currents with relatively small step changes, it is expected to achieve more accurate test results, especially for the PG faults; however, this will increase the time required for the simulations and preparation of the training set.

Based on the conditions mentioned in Section III-A, regarding each fault type, there are a total number of 100 and 150 test patterns for each fault resistance and prefault current, respectively. The fault-location results for the different fault resistance and prefault current values are presented in Tables III and IV, respectively. As can be seen from these tables, the accuracy of the fault-location method remains in an acceptable range for the different values of fault resistance and prefault current. However, in Table III, an increase in the fault resistance has led to a decrease in the prediction accuracy, especially for the PG faults.

C. Other Similarity/Distance Measures

There are other similarity/distance measures which can be considered as alternatives to the Pearson's measure. The cosine similarity is another similarity measure that can be used instead of the Pearson correlation coefficient in the proposed fault-location algorithm. This measure, which represents the similarity of two vectors' directions, can be defined as follows for the two sample vectors $x = \{x_1, x_2, \ldots, x_n\}$ and $y = \{y_1, y_2, \ldots, y_n\}$[14]:

$$c(x,y) = \frac{\sum_{i=1}^{n} x_i.y_i}{\sqrt{\sum_{i=1}^{n}(x_i)^2}\sqrt{\sum_{i=1}^{n}(y_i)^2}} \tag{5}$$

where $c(x,y)$ is the cosine similarity or the cosine of angle between the two vectors x and y and, thus, $c(x,y) \in [-1,+1]$. The greater values of c indicate higher similarity of two signals.

TABLE III
RESULTS FOR THE DIFFERENT FAULT TYPES AND FAULT RESISTANCES

	Percentage fault-location errors- PG faults				Percentage fault-location errors- PN faults				Percentage fault-location errors- PNG faults			
Fault resistance (Ohm)	Min. error (%)	Max. error (%)	Average error (%)	Fraction of errors >1%	Min. error (%)	Max. error (%)	Average error (%)	Fraction of errors >1%	Min. error (%)	Max. error (%)	Average error (%)	Fraction of errors >1%
2	0.0000	0.8850	0.2238	0%	0.0000	0.4700	0.0653	0%	0.0000	0.3300	0.0677	0%
15	0.0000	1.3900	0.2668	1%	0.0000	0.6700	0.0788	0%	0.0000	0.6700	0.0781	0%
20	0.0000	1.1150	0.2957	1%	0.0000	0.4300	0.0639	0%	0.0000	0.5700	0.0658	0%
40	0.0000	1.2150	0.3112	2%	0.0000	0.4300	0.0696	0%	0.0000	0.4800	0.0631	0%
60	0.0000	1.1700	0.3003	1%	0.0000	0.7701	0.0885	0%	0.0000	0.4700	0.0651	0%
80	0.0000	1.6050	0.3310	8%	0.0000	1.0300	0.1004	1%	0.0000	0.5700	0.0961	0%

TABLE IV
RESULTS FOR THE DIFFERENT FAULT TYPES AND PREFAULT CUIRRENTS

	Percentage fault-location errors- PG faults				Percentage fault-location errors- PN faults				Percentage fault-location errors- PNG faults			
Pre-fault current (A)	Min. error (%)	Max. error (%)	Average error (%)	Fraction of errors >1%	Min. error (%)	Max. error (%)	Average error (%)	Fraction of errors >1%	Min. error (%)	Max. error (%)	Average error (%)	Fraction of errors >1%
800	0.0000	1.2150	0.3469	3.3333%	0.0000	0.4300	0.0668	0%	0.0000	0.4700	0.0644	0%
1000	0.0000	1.6050	0.2895	4.0000%	0.0000	1.0300	0.0645	0.6667%	0.0000	0.5700	0.0662	0%
1400	0.0000	0.8800	0.2394	0%	0.0000	0.7701	0.0972	0%	0.0000	0.5500	0.0795	0%
1600	0.0000	1.3500	0.2767	1.3333%	0.0000	0.6700	0.0824	0%	0.0000	0.6700	0.0805	0%

In order to convert the similarity measure c to a positive valued distance metric, a procedure similar to (2) can be applied

$$d_c(x,y) = 1 - c(x,y) \tag{6}$$

where $d_c(x,y)$ is the cosine distance of the two signals x and y, and $d_c(x,y) \in [0,2]$.

Now, in the proposed fault-location algorithm, the cosine similarity and the cosine distance are used instead of the Pearson correlation coefficient and the Pearson distance, respectively. In this case, the obtained average fault-location errors for the PG, PN, and PNG faults are 0.8682%, 0.1041%, and 0.0830%, respectively. It can be observed that compared to the case of using the Pearson correlation coefficient, the accuracy of fault-location estimation has decreased, especially for the PG faults.

Another commonly used distance measure is the Euclidean metric. The Euclidean distance of the two signals $x = \{x_1, x_2, \ldots, x_n\}$ and $y = \{y_1, y_2, \ldots, y_n\}$, $d_e(x,y)$ is calculated as follows:

$$d_e(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}. \tag{7}$$

Considering the interpretation that smaller distance values correspond to higher similarity of vectors, the Euclidean distance is used in the fault-location algorithm. It is worth noting that in order to have smaller numerical values for Euclidean distances, per-unit values of voltage samples are considered (with the reference base voltage of 500 kV). In this case, the obtained average fault-location errors for the PG, PN, and PNG faults are 1.0452%, 0.1053%, and 0.0842%, respectively. It can be observed that compared to the case of using the Pearson correlation coefficient as well as the case of using the cosine similarity, the accuracy of fault-location estimation has decreased, especially for the PG faults.

D. Current Instead of Voltage

In this section, instead of the voltage signal, the current signal is used in the fault-location algorithm and, thus, the training and test patterns are regenerated. However, the voltage signal is used for determining the beginning moment of the 10-ms data window of the postfault current signal as well. In this case, the average fault-location error for the PG, PN, and PNG faults is 1.4747%, 0.5270%, and 0.4334%, respectively. It can be observed that in the case of using the current signal instead of the voltage signal, the accuracy of fault locating has decreased.

As can be observed in Fig. 3, immediately after the fault occurrence in the transmission line, the current signal faces a momentary overshoot, and then as a result of the control response, goes to the minimum value with some oscillations due to the oscillations of the voltage signal [12]. The momentary overshoot and the control response make the postfault current signal shapes relatively similar in various fault locations; however, there is still a little discriminant dissimilarity, but it is not enough for accurate fault locating.

E. Effect of Sampling Frequency

Here, the effect of lower sampling frequencies is studied through downsampling the existing voltage signals. Note that using lower sampling frequencies in a fixed length of the time window will reduce the number of features in the patterns. The fault-location algorithm is applied on the voltage signals with various sampling frequencies, and the average percentage errors for the different sampling frequencies are presented in Table V. As can be seen from this table, the dependency of the method's accuracy on the sampling frequency is not considerable. In fact, the sampling frequency can be set to a value as low as 5 kHz with a little loss of accuracy.

F. Effect of Noisy Measurements

Here, the noise impact on the proposed algorithm is investigated. In this regard, the voltage samples used in the test pat-

TABLE V
PERCENTAGE ERRORS FOR THE DIFFERENT FAULT TYPES AND SAMPLING FREQUENCIES

Fault Type	Sampling Frequency (kHz)					
	80	40	20	10	5	2.5
PG	0.2881	0.2939	0.2967	0.3020	0.3151	0.3909
PN	0.0777	0.0909	0.1147	0.1322	0.1690	0.2364
PNG	0.0726	0.0827	0.0985	0.1164	0.1510	0.2014

TABLE VI
PERCENTAGE ERRORS FOR THE DIFFERENT FAULT TYPES AND ADDITIVE NOISE LEVELS

Fault Type	SNR (dB)					
	∞	30	25	20	15	10
PG	0.2881	0.2917	0.2974	0.3215	0.3932	0.5477
PN	0.0777	0.0782	0.0791	0.0843	0.0928	0.1074
PNG	0.0726	0.0738	0.0751	0.0769	0.0829	0.1019

terns are corrupted by white Gaussian noise in different levels, while the training patterns remain as before. Then, the fault-location algorithm is applied on the noisy test patterns. The average percentage errors in fault-location estimation for test patterns with different signal-to-noise ratios (SNRs) are presented in Table VI. As can be observed from this table, the proposed method is relatively robust to the additive noise. However, considering a high level of additive noise, such as the one with the SNR of 10 dB, the accuracy may degrade to some extent compared to the noiseless case (infinite SNR).

G. Effect of Length of the Time Window

The length of the time window for extracting the voltage samples in the fault-location algorithm has been set to 10 ms. Note that by considering wider time windows in a specific sampling frequency, the dimension of feature vectors will increase. In this section, the proposed method is applied to the generated patterns considering various time window lengths, and the average errors for the different data windows are illustrated in Fig. 5. As can be seen from this figure, the average estimation error for the PG faults has increased with increasing the time window length, and the lowest average error belongs to the 5-ms time window. On the other hand, for the PN and PNG faults, the average estimation errors have their highest value in the 5-ms time window, but in the wider time windows, the average errors are almost constant in a relatively lower value. According to the results of Fig. 5, the adoption of the 10-ms data window seems reasonable. However, it is also possible to consider different time window lengths for different fault types.

It is worth noticing that the variations of estimation error versus the length of the time window can be different depending on the considered conditions for patterns generation or the used similarity measure. For instance, the average errors for the different data windows in the case of using the cosine similarity measure are illustrated in Fig. 6. As can be seen from this figure, the 20-ms data window is suitable for all fault types; however, even with this length of time window, the prediction accuracy is lower compared to the case of using the Pearson correlation coefficient.

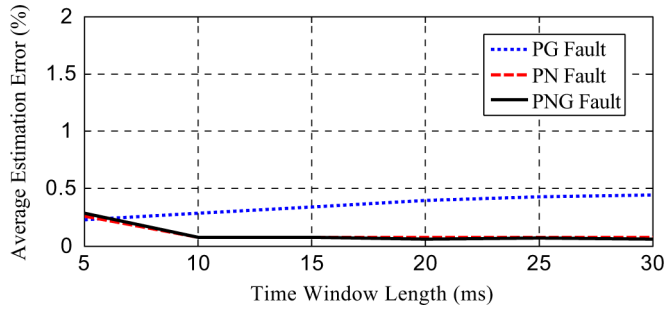


Fig. 5. Average percentage errors for the different fault types and time window lengths (using Pearson's measure).

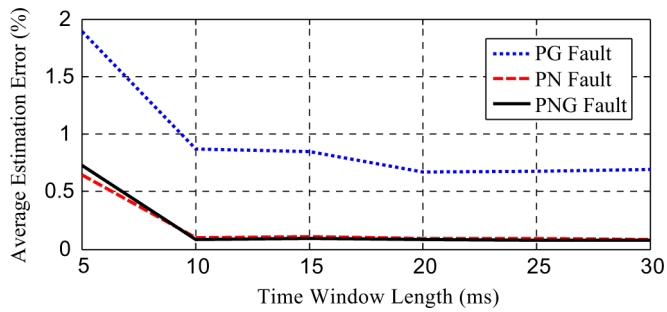


Fig. 6. Average percentage errors for the different fault types and time window lengths (using the cosine similarity measure).

H. Comparison With Other Works

Unlike the proposed method, the traveling-wave-based fault-location methods generally require a high sampling frequency of several megahertz to obtain the desired accuracy, and as shown in [3] and [7], the accuracy of them may degrade significantly with decreasing the sampling frequency to about 100 kHz. In [5], using a traveling-wave-based method and by adoption of the sampling frequency of 80 kHz, the location estimation of only five faults with zero fault resistance was associated with the average percentage error of 0.75%.

The proposed approach utilizes only the single-ended voltage measurements. Most of the accurate traveling-wave-based fault-location methods are founded on the double-ended measurements. The nontraveling-wave fault-location principle of [8] is based on the synchronized double-ended voltage and current measurements as well. Fault locating with the approach of [8], considering different fault resistances and five fault locations along a $\pm$500 kV, 1000-km-long bipolar HVDC transmission line, was associated with the average estimation errors of 0.2131%, 0.3043%, and 0.3507% for the PG, PN, and PNG faults, respectively. It can be comprehended that the approach of [8] led to higher percentage errors in location estimation of the PN and PNG faults compared to those obtained by the proposed method, 0.0777% for the PN faults and 0.0726% for the PNG faults. However, in the case of PG faults, the average percentage error obtained by the approach of [8] is slightly lower compared to that obtained by the proposed method, 0.2881%. Of course, this can be acceptable with attention to less required measurement data and no need for transmitting and synchronizing measurements of both ends in the proposed method.

IV. Conclusion

In this paper, a solution for the fault-location problem in HVDC transmission lines, from the machine-learning and the pattern-recognition point of views has been presented. In the proposed fault-location method, which is based on the similarity measure of the postfault signals using the Pearson correlation coefficient, only the voltage samples measured at one of the line terminals are required. Therefore, in addition to increasing the reliability in access to the required measured data, problems such as transmitting and synchronizing measurements of both ends, and the combination of measurement errors which might occur when more than one signal is used, are not relevant. The results of numerical studies performed on a bipolar HVDC system indicate sufficient accuracy and efficiency of the proposed method. For the primarily generated test patterns, the obtained average errors for the PG, PN, and PNG faults are 0.2881%, 0.0777% and 0.0726%, respectively. The impact of noise and the effect of using alternative similarity/distance measures, current signals instead of voltage signals, different sampling frequencies, and different time window lengths have been investigated in the sample system. The examinations show that unlike the traveling-wave-based methods, the dependency of the method's accuracy on the sampling frequency is not significant.

References

[1] O. M. K. K. Nanayakkara, A. D. Rajapakse, and R. Wachal, "Location of DC line faults in conventional HVDC systems with segments of cables and overhead lines using terminal measurements," *IEEE Trans. Power Del.*, vol. 27, no. 1, pp. 279–288, Jan. 2012.

[2] O. M. K. K. Nanayakkara, A. D. Rajapakse, and R. Wachal, "Traveling-wave-based line fault location in star-connected multiterminal HVDC systems," *IEEE Trans. Power Del.*, vol. 27, no. 4, pp. 2286–2294, Oct. 2012.

[3] M. B. Dewe, S. Sankar, and J. Arrillaga, "The application of satellite time references to HVDC fault location," *IEEE Trans. Power Del.*, vol. 8, no. 3, pp. 1295–1302, Jul. 1993.

[4] C. Ping, X. Bingyin, L. Jing, and G. Yaozhong, "Modern travelling wave based fault location techniques for HVDC transmission lines," *Trans. Tianjin Univ.*, vol. 14, no. 2, pp. 139–143, Apr. 2008.

[5] A. Swetha, P. K. Murthy, N. Sujatha, and Y. Kiran, "A novel technique for the location of fault on a HVDC transmission line," *J. Eng. Appl. Sci.*, vol. 6, no. 11, pp. 62–67, Nov. 2011.

[6] M. Ando, E. O. Schweitzer, and R. A. Baker, "Development and field-data evaluation of single-end fault locator for two-terminal HVDC transmission lines-part 2: algorithm and evaluation," *IEEE Trans. Power App. Syst.*, vol. PAS-104, no. 12, pp. 3531–3537, Dec. 1985.

[7] Y.-J. Kwon, S.-H. Kang, D.-G. Lee, and H.-K. Kim, "Fault location algorithm based on cross correlation method for HVDC cable lines," in *Proc. IET Int. Conf. Develop. Power Syst. Protect.*, Glasgow, U.K., 2008, pp. 360–364.

[8] J. Suonan, S. Gao, G. Song, Z. Jiao, and X. Kang, "A novel fault-location method for HVDC transmission lines," *IEEE Trans. Power Del.*, vol. 25, no. 2, pp. 1203–1209, Apr. 2010.

[9] *"PSCAD/EMTDC User's Guide,"* Manitoba HVDC Research Ctr., Winnipeg, MB, Canada, 2005.

[10] M. Szechtman, T. Margaard, J. P. Bowles, C. V. Thio, D. Woodford, T. Wess, R. Joetten, G. Liss, M. Rashwan, P. C. Krishnayya, P. Pavlinec, V. Kovalev, K. Maier, J. Gleadow, J. L. Haddock, N. Kaul, R. Bunch, R. Johnson, G. Dellepiane, and N. Vovos, "The CIGRE HVDC benchmark model-a new proposal with revised parameters," *Electra*, no. 157, pp. 61–65, Dec. 1994.

[11] The lower Churchill project DC1010-voltage and conductor optimization newfoundland and labrador hydro-lower Churchill project, Exhibit CE-01 Rev.1 (Public) Apr. 2008. [Online]. Available: http://www.pub.nf.ca/applications/muskratfalls2011/files/exhibits/abridged/CE-01R1-Public.pdf

[12] M. O. Faruque, Y. Zhang, and V. Dinavahi, "Detailed modeling of cigre HVDC benchmark system using PSCAD/EMTDC and PSB/SIMULINK," *IEEE Trans. Power Del.*, vol. 21, no. 1, pp. 378–387, Jan. 2006.

[13] G. W. Swift, "The spectra of fault-induced transients," *IEEE Trans. Power App. Syst.*, vol. PAS-98, no. 3, pp. 940–947, May/Jun. 1979.

[14] M. Grcar, D. Mladenic, B. Fortuna, and M. Grobelnik, "Data sparsity issues in the collaborative filtering framework," in *Advances in Web Mining and Web Usage Anal., Lecture Notes in Comput. Sci*, O. Nasraoui, O. Zaiane, M. Spiliopoulou, B. Mobasher, B. Mas, and P. S. Yu, Eds. New York: Springer, 2006, vol. 4198, pp. 58–76.

[15] D. Naidoo and N. M. Ijumba, "HVDC line protection for the proposed future HVDC systems," in *Proc. Int. Conf. Power Syst. Technol.—PowerCon*, Singapore, 2004, vol. 2, pp. 1327–1332.

[16] Y. Zhang, N. Tai, and B. Xu, "Fault analysis and traveling-wave protection scheme for bipolar HVDC lines," *IEEE Trans. Power Del.*, vol. 27, no. 3, pp. 1583–1591, Jul. 2012.

[17] Z. Xiao-Dong, T. Neng-Ling, J. S. Thorp, and Y. Guang-Liang, "A transient harmonic current protection scheme for HVDC transmission line," *IEEE Trans. Power Del.*, vol. 27, no. 4, pp. 2278–2285, Oct. 2012.

[18] T. V. Prasad and S. I. Ahson, "Data mining for bioinformatics-microarray data," in *Bioinformatics: Applicat. in Life and Environmental Sci*, M. H. Fulekar, Ed. New York: Springer, 2009, pp. 77–144.

[19] C. G. Atkeson, A. W. Moore, and S. Schaal, "Locally weighted learning," *Artif. Intell. Rev.*, vol. 11, no. 1, pp. 11–73, Feb. 1997.

[20] *"MATLAB User's Guide: R2012a Documentation,"* MathWorks Inc, Natick, MA, 2012.

Mohammad Farshad (S'11) 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 interest is the application of intelligent systems in 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.