Determining the Fault Position in the Power System

*Note: Sub-titles are not captured in Xplore and should not be used

Mario Šipoš
Croatian Armed Forces
Zagreb, Croatia
mario.sipos86@gmail.com

Zvonimir Klaić
Department of Power Engineering
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek
Osijek, Croatia
zvonimir.klaic@ferit.hr

Emmanuel Karlo Nyarko
Chair of Automation and Robotics
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek
Osijek, Croatia
karlo.nyarko@ferit.hr

Krešimir Fekete
Department of Power Engineering
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek
Osijek, Croatia
kresimir.fekete@ferit.hr

Mario Primorac
Department of Power Engineering
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek
Osijek, Croatia
mario.primorac@ferit.hr

Damir.Sljivac
Department of Power Engineering
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek
Osijek, Croatia
damir.sljivac@ferit.hr

***Abstract*— This paper describes the determination of the fault location in the power system. This problem is solved in two steps: by determining the line on which the fault occurred using the classification algorithm of the decision tree and determining the fault location on the line by linear regression. In the first step, the decision tree algorithm is used to classify the short circuit and to determine the faulted line based on the information about the residual voltages (voltage dips) on the buses due to the occurrence of a short circuit. In the second step, linear regression is used to find the fault location on the faulted line. The algorithm is tested using the 39-bus IEEE test system.**

Keywords— Fault position, decision tree, linear regression, voltage dips

I. Introduction

Classification is one of the major problems in machine learning. A prediction model called the decision tree is a widely used method for solving this problem. The decision tree method [1], [2] can be represented using a graph, where the classical construction of the decision tree is implemented using statistical methods. Due to the simple and effective classification process, that has become a popular method used for classification. It has simplified the process of building a set of classifiers: for example, deciduous forests, which are defined as sets of specific decision trees, which improves the quality of the classification. Also, this method is an effective supervised technique in high-dimensional data [3]. It has been developed in power systems for high impedance fault detection [4], transient stability prediction [5], where overview on applications of data mining in power systems is given in [6]. Decision tree method was used for fault detection and classification in solar photovoltaic array in [7]. Fault type is describes in Single-Circuit Transmission Lines [8], and fault zone is presented in identification and fault classification in AC transmissions line [9].

The process of creating decision trees is a complex problem of multi-criteria decisions that relate to the rules used in optimal data separation. Decision trees can easily represent complex terms whose definitions can be described based on feature space. It has very competitive with other classification algorithms with respect to memory usage.

In this paper decision tree method is used to determine the faulted line in the power system based on the voltage dips information. After the faulted line is found, linear regression method is used to find location of the fault. To test and verify the proposed method, IEEE 39 bus system is used. In second section, determination of the faulted line using decision tree is described. Determination of fault position using linear regression method is given in section III. Conclusion is given in the last chapter.

II. Determining the Faulted Line

A set of samples is obtained in a controlled classification. This set consists of n observations, also called objects or samples.

$$X = \{x_1, x_2, \dots, x_i, \dots x_n\} \tag{1}$$

Each of the observations x_i is described by the attributes m, which are also called the features $a_1, a_2, \dots, a_m$. With $a_j \in A_j$, j=1,…, m, where A_j denotes the domain of the j^{th} attribute. In this way $a_1, a_2, \dots, a_m$ form the feature space $A_1 \times A_2 \times A_m$. The values of these attributes can be quantitative (e.g., price) or categorical (for example, weather: “rainy” or “sunny”). Each observation belongs to one of C different and known decision classes, which can be represented by an expression.

$$x_i = (V_i, c_i), v^j_i \in A_j, c_i \in \{1, \dots, C\} \tag{2}$$

978-1-7281-9759-3/20/$31.00©2020 European Union

Where is $V_i=[v_i^1, \ldots v_i^m]$ vector in the m-dimensional feature space, v^j_i is the value of the attribute A_j for observation (object) x_i and c_i is the designation of the class (also called the decision class) of the object observation x_i. Therefore, X can be given by the expression.

$$X : \{(V_i, c_i)\}^n{}_{i=1} \tag{3}$$

Accordingly, the classification problem can be defined as the determining of how to assign an object to a class, knowing that there are different classes of decision C and that each object belongs to one of them.

The decision tree is a classifier that makes decision rules, commonly used in operational research, especially in decision analysis, to determine the optimal strategy for achieving a goal. The decision tree is an acyclic, directed graph in which all the vertices are called nodes, the edges are called branches, the vertices without descendants are leaves, and the root is the only vertex without parent, Figure 1. All nodes contain attribute tests A, which are generated according to the selected division criteria. They represent a way of specifying a data division that splits data by attribute values for that data. Finally, all test results are presented with branches [10].

The Decision Tree Algorithm was created in Matlab. 3366 three-phase short circuits were used at different line locations in the IEEE 39 bus test system (Figure 2). The data were separated so that 2346 (70%) of the data were used for training and 1020 data (30%) for testing the trained decision tree. Voltage dips that are taken into consideration are based on the PQ monitors position. In this paper, monitors are located at the certain buses according to the two optimization methods: binary bat algorithm and GLPK. There were two tests conducted in this paper. The first test used the optimal number and location of voltage dips based on binary bat algorithm and second based on GLPK (GNU Linear Programming Kit - GLPK) optimal solution. The results were analyzed by the confusion matrix. In the first test of the 1020 values used that were not in the training set and the decision tree encountered them for the first time, 58 were misclassified. In other words, 58 (5.6%) times faulted line was predicted wrong. In the second test, the results indicated 60 (60/1020 = 5.88%). At first glance, the number of misclassified lines appears to have increased by two due to the smaller number of measuring devices, however, by comparing the confusion matrices they differ in certain cases shown in Table 1.

From Table 1. it can be seen that the difference in the errors occurred at lines 10, 11, 16, 19, 25 and 34. The other errors are the same with both methods. From this example, it can be concluded that the arrangement of voltage dip measuring devices not only affects the number of misclassifications determined by the decision tree method but also affects the accuracy of the classification. Different voltage dip meter locations will result in different classification errors.

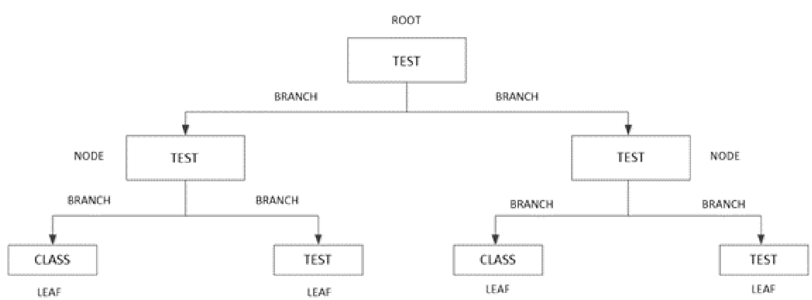

Fig. 1. Decision tree

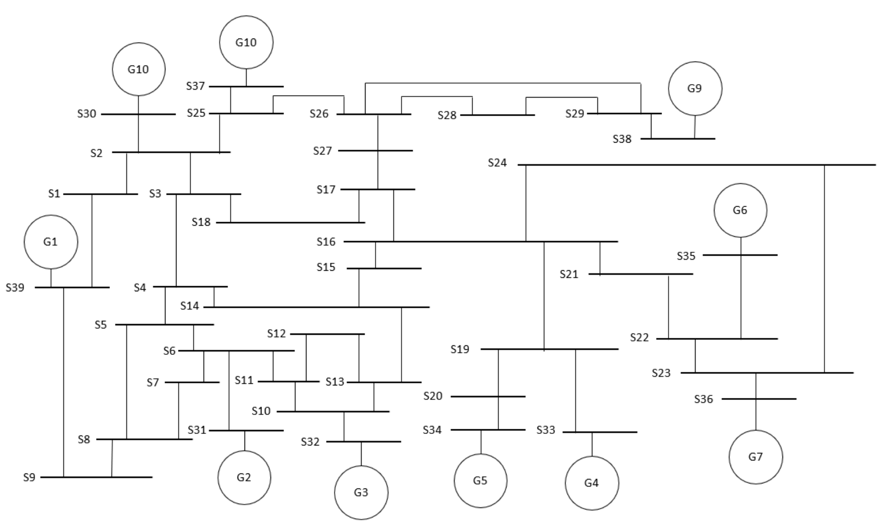

Fig. 2. IEEE 39-bus test system

TABLE I. COMPARISON OF TESTS

Binary bat algorithm			GLPK		
Predicted (Line)		**Real (Line)**	**Predicted (Line)**		**Real (Line)**
Line 10	30	30	Line 10	29	30
			Line 11	1	
Line 11	29	30	Line 11	27	30
Line 9	1		Line 10	3	
Line 16	29	30	Line 16	30	30
Line 17	1				
Line 19	28	30	Line 19	28	30
Line 20	1		Line 20	1	
Line 8	1		Line 18	1	
Line 25	29	30	Line 25	30	30
Line 21	1				
Line 34	24	30	Line 34	23	30
Line 22	6		Line 33	7	

III. DETERMINING THE FAULT POSITION ON THE LINE

Regression analysis test the dependence of one variable on another or more other variables using different methods of testing for dependence. Dependent variables are those whose variations are explained by other variables which are called independent variables. A regression model is an algebraic model that analytically expresses the statistical relationship between phenomena and is the basis of any analysis. Multiple linear regression is the basic regression model and is used to represent analytically the covariance of one numerical (dependent) variable using several other numerical (independent) variables. A general model of multiple linear regression is given by the expression:

$$Y=\alpha+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_j X_j+\cdots+\beta_K X_K+e \qquad (4)$$

where Y is the dependent variable, are X_1, X_2, …, X_K are values of the independent variable, α and β_j, j = 1,2, …, K values of unknown parameters, e is random variable (relation error) [11].

Multiple linear regression code was created in Matlab, the line to be tested needs to be selected. The 30 fault points shown in the table were used and the estimated fault percentage of the line is below. Table 2. gives a comparison of 39, 14, and 13 voltage dip devices.

From Table 2. it can be seen that when using data from 13 voltage dip measuring devices the largest error is 1.08% when predicting fault at 4% of line 8 length. As the length of the line is 51.18 km, 1% of the length of the line is 0.5118 km, which equals 511.8 m. However, when predicting fault at 18% of line 8 length, the absolute error is 0.65% which is less than 0.96% at 39 voltage dip meters but more 0.47% which is achieved when are use 14 voltage dip meters.

Table 3. shows the results of the fault prediction on the percent of the line for the longest Line 26-29 (Line 33), which is 247.96 km long. Where "AE" in the table represents an absolute error between the actual value and the predicted value and No.MD represents the number of measuring devices installed. From Table 3. it can be seen that the largest absolute error occurs when predicting fault at 17% of Line 33 length and is 3.39%, using the optimum location of voltage dip meter devices determined by the binary bat algorithm.

As the line length is 247.96 km, 3.39% of the line length is 8.4 km. The largest absolute error in the use of measuring devices on all buses is 2.12% in predicting the fault location at 66% of the line length, where the absolute error with binary bat algorithm (14 inputs from measuring devices) is 0.64% or less than 1%. Using the optimum location of devices for measuring voltage dip according to GLPK, the highest absolute error is achieved when predicting fault at 79% of line length and is 2.17%, while for predicting the same fault location, using 14 measuring devices, the absolute error is 1.62 %.

TABLE II. COMPARISON OF 39, 14, AND 13 VOLTAGE DIP DEVICES.

No.MD	2	4	10	11	14	15	17	18	28	31
39	1,9	3,24	9,86	11,36	13,55	14,81	17,17	18,96	28,5	31,12
AE	0,1	0,76	0,14	0,36	0,45	0,19	0,17	0,96	0,5	0,12
14	1,07	3,11	9,66	10,61	14,51	15,19	17,49	18,47	28,43	31,74
AE	0,93	0,89	0,34	0,39	0,51	0,19	0,49	0,47	0,43	0,74
13	1,11	2,92	9,68	10,55	14,53	15,22	17,38	18,65	28,63	31,70
AE	0,89	1,08	0,32	0,45	0,53	0,22	0,38	0,65	0,63	0,70

Line 8 fault location in %										
No.MD	34	35	37	43	47	48	55	60	61	66
39	34,58	35,1	37,07	43,72	47,16	48,64	55,96	59,54	60,94	65,21
AE	0,58	0,1	0,07	0,72	0,16	0,64	0,96	0,46	0,06	0,79
14	34,33	35,23	37,00	43,12	47,39	47,63	54,74	59,91	60,72	65,81
AE	0,33	0,23	0,00	0,12	0,39	0,37	0,26	0,09	0,28	0,19
13	34,29	35,26	36,98	42,92	47,29	47,65	54,82	59,69	60,64	65,68
AE	0,29	0,26	0,02	0,08	0,29	0,35	0,18	0,31	0,36	0,32
Line 8 fault location in %										
No.MD	67	68	75	83	85	86	87	91	94	98
39	67,05	68	74,39	82,94	85,34	86,06	87,2	91,08	94,23	98,08
AE	0,05	0	0,61	0,06	0,34	0,06	0,2	0,08	0,23	0,08
14	66,64	67,77	74,30	82,35	84,92	85,71	86,81	91,00	94,55	98,84
AE	0,36	0,23	0,70	0,65	0,08	0,29	0,19	0,00	0,55	0,84
13	66,77	67,71	74,20	82,40	84,84	85,71	86,75	91,26	94,40	98,72
AE	0,23	0,29	0,80	0,60	0,16	0,29	0,25	0,26	0,40	0,72

TABLE III. PREDICTED VALUES FOR LINE 33 (LINE 26-29).

No.MD	2	4	5	13	17	19	22	23	28	31
39	2,19	2,47	4,69	12,19	17,48	18,19	22,26	22,52	28,58	32,07
AE	0,19	1,53	0,31	0,81	0,48	0,81	0,26	0,48	0,58	1,07
14	1,01	2,47	3,40	13,41	20,39	19,27	20,34	23,18	29,66	32,88
AE	0,99	1,53	1,60	0,41	3,39	0,27	1,66	0,18	1,66	1,88
13	2,57	4,08	3,88	12,68	17,86	18,62	21,19	22,51	28,24	31,96
AE	0,57	0,08	1,12	0,32	0,86	0,38	0,81	0,49	0,24	0,96
Line 33 fault location in %										
No.MD	36	39	41	42	45	47	48	49	54	58
39	37,43	39,43	41,23	41,72	45,74	47,25	48,88	48,62	53,81	57,59
AE	1,43	0,43	0,23	0,28	0,74	0,25	0,88	0,38	0,19	0,41
14	36,09	40,55	41,08	41,58	42,37	45,20	48,34	49,77	52,79	58,94
AE	0,09	1,55	0,08	0,42	2,63	1,80	0,34	0,77	1,21	0,94
13	36,96	40,55	40,62	41,66	44,40	46,89	47,74	49,67	53,46	58,35
AE	0,96	1,55	0,38	0,34	0,60	0,11	0,26	0,67	0,54	0,35
Line 33 fault location in %										
No.MD	61	62	64	66	69	70	79	83	84	96
39	60,23	61,07	63,58	63,88	68,52	70,20	78,56	82,66	84,19	95,80
AE	0,77	0,93	0,42	2,12	0,48	0,20	0,44	0,34	0,19	0,20
14	60,49	60,39	61,62	65,36	67,76	68,32	77,38	82,84	82,85	97,59
AE	0,51	1,61	2,38	0,64	1,24	1,68	1,62	0,16	1,15	1,59
13	61,07	62,52	63,70	66,53	69,27	69,02	76,83	83,49	83,03	97,24
AE	0,07	0,52	0,30	0,53	0,27	0,98	2,17	0,49	0,97	1,24

IV. CONCLUSION

The classification of fault-affected lines was performed using the decision tree algorithm, where the fault line was determined with great precision. Out of 1020 test samples, 58 (5.6%) times faulty line was classified wrong, which gives an accurate method of classifying fault affected line. After determining the faulted line, a multiple linear regression predicts the fault location at the line. When testing for Line 8, 5.18 km long, the error was less than 1%, which is 511.8 m (1% of line length), however, for Line 33, 247.96 km long, the error was 3.39% of line length, which equals 8.4 km. So, using multiple linear regression to determine the fault location at the line length will not give satisfactory results for all lines.

REFERENCES

[1] L. Rokach and O. Maimon, "Top-down induction of decision trees classifiers - A survey," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 35, no. 4, pp. 476–487, 2005.

[2] R. C. Barros, A. C. P. L. F. De Carvalho, and A. A. Freitas, *SPRINGER BRIEFS IN COMPUTER SCIENCE Automatic Design of Decision-Tree Induction Algorithms*. 2015.

[3] Jiawei Han Micheline Kamber, *DataMining Concepts and Techniques*. 2006.

[4] Y. Sheng and S. M. Rovnyak, "Decision tree-based methodology for high impedance fault detection," *IEEE Transactions on Power Delivery*, vol. 19, no. 2, pp. 533–536, 2004.

[5] S. Rovnyak, J. Thorp, S. Kretsinger, and D. Brown, "Decision Trees for Real-Time Transient Stability Prediction," *IEEE Transactions on Power Systems*, vol. 9, no. 3, pp. 1417–1426, 1994.

[6] H. Mori, "State-of-the-art overview on data mining in power systems," *2006 IEEE PES Power Systems Conference and Exposition, PSCE 2006 - Proceedings*, pp. 33–37, 2006.

[7] Y. Zhao, L. Yang, B. Lehman, J. F. De Palma, J. Mosesian, and R. Lyons, "Decision tree-based fault detection and classification in solar photovoltaic arrays," *Conference Proceedings - IEEE Applied Power Electronics Conference and Exposition - APEC*, pp. 93–99, 2012.

[8] A. Jamehbozorg and S. M. Shahrtash, "A decision tree-based method for fault classification in double-circuit transmission lines," *IEEE Transactions on Power Delivery*, vol. 25, no. 4, pp. 2184–2189, 2010.

[9] S. R. Samantaray, "Decision tree-based fault zone identification and fault classification in flexible AC transmissions-based transmission line," *IET Generation, Transmission and Distribution*, vol. 3, no. 5, pp. 425–436, 2009.

[10] J. Kozak, *Decision Tree and Ensemble Learning Based on Ant Colony Optimization*, 1st ed. Katowice, Poland: Springer International Publishing, 2019.

[11] Ivan Šošić, *Primjenjena statistika*, 2nd ed. Školska knjiga, 2006.