January-February 2020
ISSN: 0193-4120 Page No. 6803 - 6808

Electrical Transmission Line Fault Detection and Classification using Convolution Neural Networks and Support Vector Machine

[1]D. Baskar, [2]P. Selvam
[1]Research Scholar, [2]Professor & Head, Department of Electrical & Electronics Engineering,
Vinayaka Mission's Kirupananda Variyar Engineering College,
Vinayaka Mission's Research Foundation (Deemed to be University), Salem, Tamil Nadu, India

Article Info
Volume 82
Page Number: 6803 - 6808
Publication Issue:
January-February 2020

Article History
***Article Received**: 18 May 2019*
***Revised**: 14 July 2019*
***Accepted**: 22 December 2019*
***Publication**: 01 February 2020*

Abstract
The detection and classification of mistaken conditions in electrical power transmission and distribution systems is a task of importance for trustworthy activity. In recent years, methods based on performance in relay security and performance of electrical components has become familiar. Moreover, they have fault in dealing with power system insufficiency. This paper proposes a new technique for the prediction of data-based transmission line faults in power systems using Convolution Neural Networks (CNN) networks and support vector machine (SVM). The chronological characteristics of multi-sourced data were taken into custody with CNN networks, which achieve well in extorting the features of time series for a long-time period. The powerful learning and mining capability of CNN networks is appropriate in power transmission and distribution for a heavy quantity of time series. SVM is introduced for classification in order to obtain the final predictive results, with a strapping generalization capability and sturdiness. The method tested on the VSB dataset and practical detection performance is attained.

***Keywords:** Power System, Fault Detection, Convolution Neural Networks (CNN), Support vector machine (SVM).*

1. Introduction

Medium voltage overhead power lines run for several hundreds of kilometers to provide power to various places in the world. These long distances make it high-priced to physically examine the lines for damage that doesn't instantaneously guide to a power outage, such as a plant twig hitting the line or a fault in the insulator. These methods of damage lead to an observable fact known as partial discharge — an electrical discharge that does not link the electrodes between an insulation system entirely. Partial discharges gradually break the power line, so left unrepaired they will eventually lead to a power outage or start a fire.

2. Problem statement

Line trip is a frequent fault that will lead to an immense power outage. In latest years, relay protection performance but the electrical component performance had been used for fault identification. The importance of the fault should be evaluated for applicability and stability. Ageing and damage to sharing equipment, poor insulation, weather changes and so on are the few important causes for line trip failures. When the line trip faults happen there is a plodding method of distribution line resistance. The electrical metrics will transform during the progression according to certain rules that include current, voltage, active power, and user reactive power.

3. Proposed Approach

The proposed approach is aimed at capturing the characteristics of this method to detect faults. Consequently, the importance of the fault within the fault documents and electrical measurements has to be mined. Assume that P is the result of fault prediction, where P = 1 represents a fault, and P = 0 represents normal operations, which is the label in network training for each sample. Figure 3.1 shows a simple flow chart for clarifying the

January-February 2020
ISSN: 0193-4120 Page No. 6803 - 6808

issue in this paper. Feature mining in power systems for big data is a considerable and complicated issue in the process. Concerning CNN networks and SVM, the key setback is to design parameters and network architectures for the prediction of high-precision faults. Distinctively, a realistic experiment is performed for real-world datasets acquired from the ENET Centre at VŠB Technical University of Ostrava in Czech Republic.

Figure 3.1: Solution Flow Chart for Fault Prediction

4. Recurrent Neural Network

Recurrent neural networks are feedforward neural networks amplified by the insertion of edges that extent adjacent time steps, launching a concept of time to the model. RNNs may not have cycles between predictable edges, like feedforward networks. On the other hand, edges that unite adjacent time phases, called recurrent edges, that form cycles, counting length cycles one that is, over time, self-connection from a node to itself.

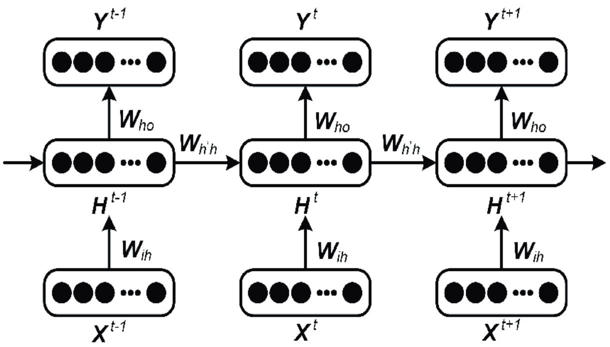

Figure 4.1: A-simple-RNN-structure-where-X-is-the-input-unit-H-is-the-hidden-unit-Y-is-the-output

An easy RNN structure is shown in the Fig. 4.1. The forward propagation process can be completed from Fig. 4.1, given by eqs. (4.1)-(3).

$$a_h^t = \sum_{i=1}^{I} w_{ih} x_i^t + \sum_{h'=1}^{H} w_{h'h} s_{h'}^{t-1} \qquad (4.1)$$

$$s_h^t = f_h\left(a_h^t\right) \qquad (4.2)$$

$$a_o^t = \sum_{h=1}^{H} w_{ho} s_h^t \qquad (4.3)$$

The notation w is the weight; a is the sum calculated via weights; f is the activation function; s is the value after calculation use the activation function; t indicates the current time of the network; i is the number of input vectors; h is the number of hidden vectors in t time; h 0 is the number of hidden vectors in t − 1 time; and o is the amount of output vectors.

5. Convolution Neural Networks (CNN)

Convolutional Neural Network (CNN) is a multi-layer neural network consisting of multiple convolution layers and alternating layers of pooling (lower sampling layer). Therefore, one or several layers of complete correlation are associated to categorize an image features created using the prior layers. CNN's amount of free parameters are significantly shortened by the use of minimal neuron link and distribution weights, which is applied to be competent over that of a fully connected network. In addition, image features have increased invariance in transformation, scaling, and deformation due to the authority of pooling layer. The basic framework of convolution neural network structures could be alienated into 5 phases: input layer, output layer, convolution layer, full connection layer, pooling layer. The subsequent segment describes the parts of the system in brief.

Input layer: convolution input layer could behave unswervingly on particular input data. The pixel value of a image was the input data for the input image.

Convolution layer: From the data of the image hypothesis, they can know that limited characteristics of a normal image might be similar or alike to another local region, that are indicates the erudite features of one region, could also be used for additional regions. For the convolution neural network, the output of the convolution layer is acquired by transforming the initial layer's filter and input feature graph (convolution kernel is constructed by 1-by-1 sliding window on feature graph), gaining a bias term and then applying it to a non-linear starting method. The attribute graph of the layer is the output rate of the convolution layer. Every filter gives a characteristic output map. Throughout this work, the input image is correspond to by X, the kth feature graph of layer i is correspond to by Ai, and the features of the kth filter of layer i are defined by the weight matrix WkiWik and the bias term bkibik. Then, the k feature graph of level i could be acquired from the below formula (5.1):

$$a_j^t = \sum_{i=1}^{I} w_{ij} x_i^t + \sum_{h=1}^{H} w_{hj} s_h^{t-1} + \sum_{c=1}^{C} w_{cj} m_c^{t-1} + b_j \qquad (5.1)$$

$$a_k^t = \sum_{i=1}^{I} w_{ik} x_i^t + \sum_{h=1}^{H} w_{hk} s_h^{t-1} + \sum_{c=1}^{C} w_{ck} m_c^{t-1} + b_k \qquad (5.2)$$

The weight sharing and local connection are the advantages of the convolution layer. The image feature may be omitted to develop during local connection, and also the communication restriction of adjacent network

January-February 2020
ISSN: 0193-4120 Page No. 6803 - 6808

layers might be shortened. To use this method the number of necessary parameters will drastically decrease.

$$a_c^t = \sum_{i=1}^{I} w_{ic} x_i^t + \sum_{h=1}^{H} w_{hc} s_h^{t-1} + b_c \qquad (5.3)$$

Pooling layer: After the image features are extorted throughout the convolution layer, the classifier may tentatively input them into the full connection layer then off the record. In the other hand, estimate is very high since the large measurement of the functions, and effortless to generate over-fitting.

While convolution layer utilizes local connection to minimize the no. of network structure associates, increasing the number of characteristic graphs would improve the feature aspect and the network is still very multifaceted, and training is still difficult. A pooling layer is typically followed by a convolution layer to further reduce the network parameters and reduce the complexity and over-fitting extent of the replica. The pooling layer is a sampling method that incorporates the output of adjacent neurons into the similar character graph. After the down-sampling mechanism has been developed, it will successfully shrink the measurement of the function and keep the efficient in order of the image, while removing redundant data and speeding up network training. The norm of this aggregation is that there is a huge similarity between the pixels in every neighboring region of the image. The region could be defined by computing the district's highest value as the sampling value, or by summing all the region's values to normal and taking the average as the sampling value.

Formula (5.4) may be used as a representation of the pooling layer activity. The neurons in this layer accept a down) (down sampling function that is used to increase or pool the feature map by norm.

$$a_l^t = \sum_{i=1}^{I} w_{il} x_i^t + \sum_{h=1}^{H} w_{hl} s_h^{t-1} + \sum_{c=1}^{C} w_{cl} m_c^t + b_l \qquad (5.4)$$

Fully connected layer: It includes various layers of full connection which is the hidden multilayer perceptron layer. The ganglion points in the posterior layer were usually linked to each point of the ganglion in the prior layer, and there is no link between the neuron nodes in the related layer. Which layer of neuron nodes promulgates forward via the weights on the line concerned, and the weights are collective for the input of the layer of neuron nodes that follows.

Output layer: The quantity of the output layer neural nodes can set as per the exact function errands. The output layer of the convolution neural network is generally a classifier when it is a classification duty.

6. Logistic Regression Classifier

The goal of the logistic regression classifiers are to learn a 0/1 classification model from the training data features by a logistic function. Assume that there is a fixed training set f($x^{(1)}$; $y^{(1)}$); ($x^{(2)}$; $y^{(2)}$); : : : ; ($x^{(m)}$; $y^{(m)}$)g, and $y^{(m)}$ 2, where m is the number of samples. The hypothesis function is given by

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1-y^{(i)}) \log(1-h_\theta(x^{(i)}))\right] \qquad (6.1)$$

where θ is the weight of the logistic regression classifier to be obtained. The result of the hypothesis function shows the probability P of y = 1, so it may be concluded that

$$P(y=0/x;\theta) = 1 - h_\theta(x) \qquad (6.2)$$

The cost function can be derived by a maximum likelihood estimate, shown in eq. (6.1).

$$P(y=1/x;\theta) = h_\theta(x) \qquad (6.3)$$

Therefore, a gradient descent method can be used to minimize the cost function.

$$d_c^t = s_k^t m_c^{t-1} + s_j^t g(a_c^t) \qquad (6.4)$$

7. Support Vector Machine

SVM is a supervised learning model for classification and regression analysis. The basic rule looks for an optimal hyperplane which is the farthest from the nearest training samples. For the linear separation in a two-dimensional plane, the classification function can be

$$f(x) = w^T x + b \qquad (7.1)$$

where *w; b* determines a straight line for classification. According to the relation of geometrical margin $\check{y}$ and functional margin $\acute{y}$ is given by

$$\tilde{\gamma} = \frac{\tilde{\gamma}}{\|\omega\|} \qquad (7.2)$$

The problem can be inferred as eq. (7.3) when the $\check{y}$ = *1*

$$max \frac{1}{\|\omega\|}, s,t., y_i(w^T x_i + b) \geq 1, i = 1, 2, \ldots, n \qquad (7.3)$$

where yi = *1* or -1 is the label of samples, and *n* is the number of samples. To widen SVM to cases in which the data is not linearly divisible, the hinge loss function is given by

$$max\left(0,1-y_i(w^T x+b)\right) \quad (7.4)$$

Then the loss function can be concluded as eq. (7.5)

$$min_{w,b}\sum_{i=1}^{N} max\left(0,1-y_i(w^T x+b)\right)+\lambda\|\omega\|^2 \quad (7.5)$$

SVM may use a kernel trick and nonlinear mapping algorithm for linear inseparability to convert samples from low-dimensional input space into high-dimensional feature space so that they can be linearly separated. SVM is generally a novel learning method with a solid theoretical base that has good robustness with less samples and efficiency of generalization in nonlinear problems related to classifiers of logistic regression. The explanation for this is that SVM's optimization target is structural risk minimization rather than empirical risk minimization. It not only guarantees the sample classification precision but also decreases the learning model dimension, corresponding to the two eq. terms (21). In short, it helps in avoiding overfitting issues. In addition, the computational problem is depend on the quantity of support vectors as an alternative to the sample space dimension, which in many ways removes the adversity dimension. Thus, SVM has limited computational complexity and excellent sturdiness.

8. Simulation Results

This section discusses the simulation condition and the outcomes of the practical experiment, which confirms that the proposed method has increased efficiency. Fig. 8.1 shows a normal sample and fault sample for the present. It is not possible to determine the difference between them easily from the figures. This is same for voltage and active control, so the hidden features would be mined for fault prediction though CNN networks. The specimens are time series with temporal information for the input of the CNN networks is translated into various dimensions.

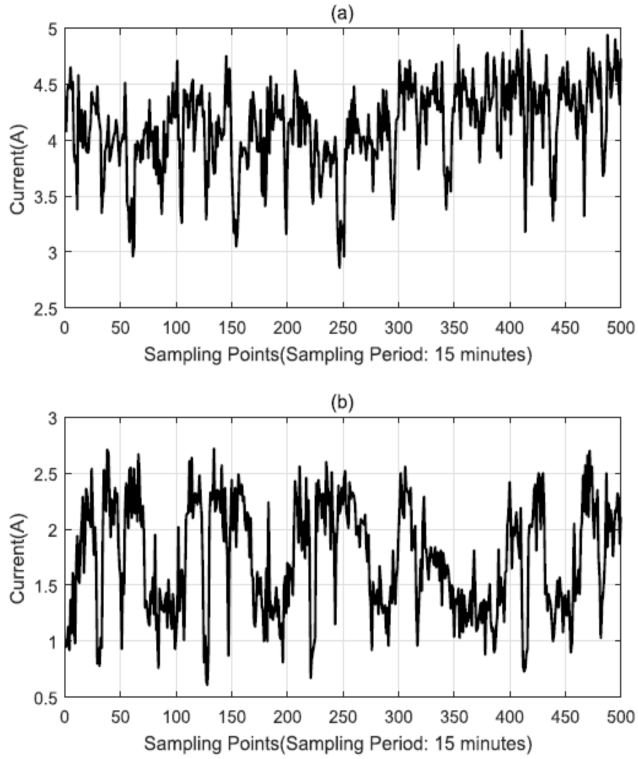

Figure 8.1: (a) Normal sample and (b) fault sample of the current variations.

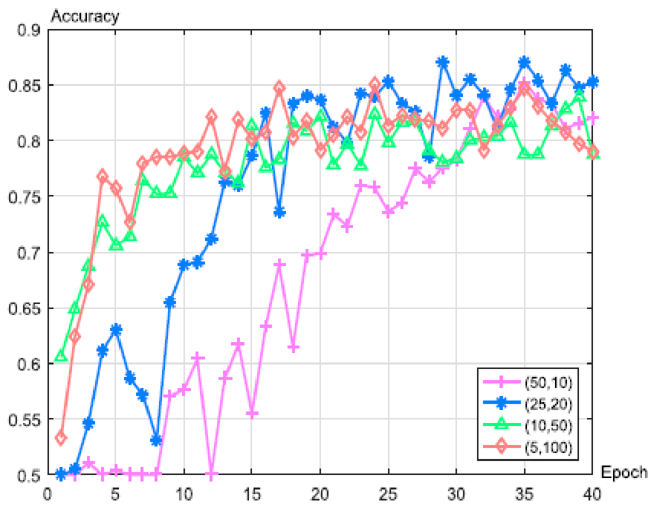

Figure 8.2: Comparative results of different input time steps and dimensions, where T in (T,D) is the time step and D in (T,D) is the dimension of the input vector.

January-February 2020
ISSN: 0193-4120 Page No. 6803 - 6808

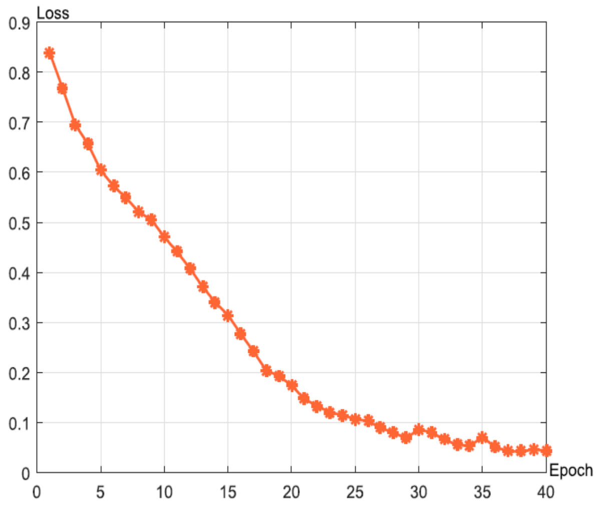

Figure 8.3: The loss in fault prediction through 40 epoches of training.

Table 1 Comparison result

Methods	Accuracy
BPNN	69.4
SAE	79.3
RNN	83.5
SAE+SVM	88.5
RNN+SVM	94.8
Proposed Methods	98.7

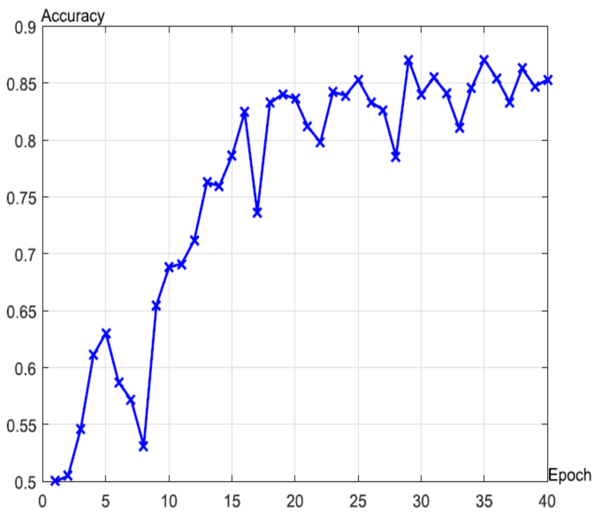

Figure 8.4: The accuracy of the fault prediction through 40 epoches of training

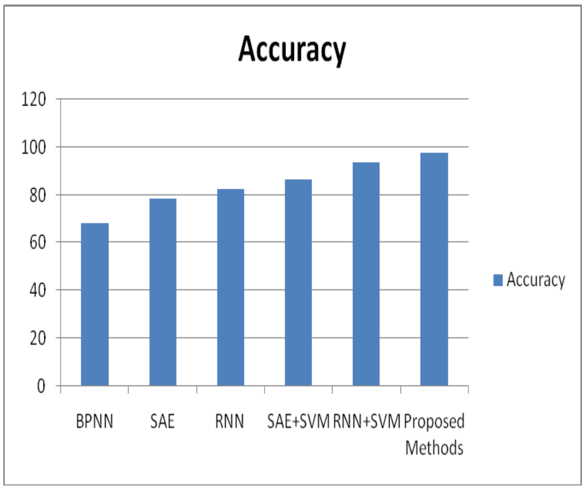

Figure 8.5: Comparison results for the proposed method

The CNN network-based fault prediction experiment is discussed in this section, Comparative results of various input time steps and dimensions, where T in (T, D) is the time phase, and D in (T, D) is the input vector dimension Fig. 8.2. The final result is the average precision of five trained models: 98.7%. Fig. 8.3 and Fig. 8.4 demonstrate the precision of the estimation of fault and failure during a training period. The precision enhances as with the training, the loss reduces. The results of comparison for the proposed method and the data-based techniques to mining are display in Fig. 8.5, where normal and fault test set precision has been involved. It displays that the growth is marked both in normal and in fault circumstances. Table 1 displays the final results.

9. Conclusion

It can be concluded that the proposed method works better than other methods like neural networks for backpropagation (BPNNs), stacked autoencoders (SAEs), RNNs, and SVM. Based on the linked hidden layer units, CNN networks can extract the temporal information from data, but BPNN and SAE do not have this capability. Compared to RNNs, CNN may use the CNN block to solve the problem of the vanishing gradient. In general, it relies on CNN networks' stronger learning capacity for time series, and SVM's good robustness and efficiency in generalization. The fault characteristics are mined at high precision from multiple measurement data sources to predict fault. Compared to approaches based on relay safety behavior and acts of electrical components, the proposed data-based approach can predict faults in a power system based on first-hand information. The suggested technique, that of transmission line trip fault prediction in power systems using CNN networks with SVM, is therefore a notable development in ensuring a power system's efficiency and stability. In summary, it is noteworthy to develop the proposed methodology as compared to the current methods to data mining. The precision of the forecast of transmission line fault will hit

January-February 2020
ISSN: 0193-4120 Page No. 6803 - 6808

around 98.7%. It is of great importance for the reliability of operation and for the stability of a power system.

References

[1] Nikita V.Tomin et al. Machine-Learning Techniques for Power-System Security Assessment. IFAC-PapersOnLine, 2016, 49(27), pp.445-450.

[2] S. Rinaldi, M. Pasetti, A. Flammini, F. De Simone, "Characterization of Energy Storage Sytems for Renewable Generators: An Experimental Testbed," IEEE International Workshop on Applied Measurements for Power Systems (AMPS), 2018.

[3] A.Muthukrishnan, J.Charles Rajesh kumar, D.Vinod Kumar, M.Kanagaraj. "Internet of image things-discrete wavelet transform and Gabor wavelet transform based image enhancement resolution technique for IoT satellite applications".57, pp.46-53, 2019.

[4] H.Kiani et.al. Coordinated transmission substations and sub-transmission networks expansion planning incorporating distributed generation, Journal of Energy, 2017, 120(1) ,pp.996-1011.

[5] N.Stalin, P.Selvam." Power Transfer efficiency Analysis of Double Intermediate-Resonator for Wireless Power Transfer". International -Journal of Advances in Engineering and Emerging Technology, 9(3), pp.130-141, 2018.

[6] D.Madeshwaran P Selvam." Back To Back Converter Based Real and Reactive Power Control with Constant Speed Operation of Dfig in Wind Mill", 8(2), pp.1855-1860, 2019.

[7] N. S. Coleman, C. Schegan, and K. N. Miu, "A study of power distribution system fault classification with machine learning techniques," 2015 North American Power Symposium (NAPS), Charlotte, NC, 2015, pp. 1-6.

[8] Jiawei Han, Micheline Kamber, DATA MINING, Concepts and Techniques (second ed.), Morgan Kaufman Publishers (2003)

[9] Hasan, Ali & Eboule, Patrick & Twala, Bhekisipho. (2017). The use of machine-learning techniques to classify power transmission line fault types and locations. 221-226.

[10] Z. Qian, Z. Y. Jie and H. Pu, "The Application of Sequential Minimal Optimization Algorithm In Short-term Load Forecasting," 2007 Chinese Control Conference, Hunan, 2007, pp. 314-317.

[11] Dongxiao Niu and Yongli Wang, "Study of the SMO algorithm based on data mining in short-term power load forecasting model," 2008 7th World Congress on Intelligent Control and Automation, Chongqing, 2008, pp. 7140-7145.

[12] Asadi Majd, A., Samet, H. & Ghanbari, T. Prot Control Mod Power Syst (2017) 2: 32.

[13] Charles Rajesh Kumar.J, Vinod Kumar.D, Baskar.D, Mary Arunsi.B, Jenova.R, M.A. Majid." VLSI design and implementation of High-performanceBinary-weighted convolutional artificial neural networks for embedded vision based Internet of Things (IoT)".16th International Learning & Technology Conference (L&T), Procedia Computer science (Elsevier), Jeddah, 2018.

[14] Manikanthan, S.V., Padmapriya, T., An efficient cluster head selection and routing in mobile WSN, International Journal of Interactive Mobile Technologies, 2019.

[15] Recioui, B. Benseghier, and H. Khalfallah, "Power system fault detection, classification and location using the K-Nearest Neighbors," 2015 4th International Conference on Electrical Engineering (ICEE), Boumerdes, 2015, pp. 1-6.