Support-vector machine and Naïve Bayes based diagnostic analytic of harmonic source identification

Mohd Hatta Jopri¹, Abdul Rahim Abdullah², Jingwei Too³, Tole Sutikno⁴, Srete Nikolovski⁵, Mustafa Manap⁶

¹¹Faculty of Electrical & Electronic Engineering Technology, Universiti Teknikal Malaysia Melaka, Malaysia
²³Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka, Malaysia
⁴Department of Electrical Engineering, Universitas Ahmad Dahlan, Indonesia
⁵Power System Department, Faculty of Electrical Engineering, University of Osijek, Croatia

Article Info

Article history:
Received Jan 24, 2020
Revised Mar 26, 2020
Accepted Apr 10, 2020

Keywords:
Harmonic source diagnosis
Naïve Bayes
S-transform
Support-vector machine

ABSTRACT

A harmonic source diagnostic analytic is a vital to identify the location and type of harmonic source in the power system. This paper introduces a comparison of machine learning (ML) algorithm which are support vector machine (SVM) and Naïve Bayes. Voltage and current features are used as the input for ML are extracted from time-frequency representation (TFR) of S-transform. Several unique cases of harmonic source location are considered, whereas harmonic voltage and harmonic current source type-load are used in the diagnosing process. To identify the best ML, the performance measurement of the propose method including accuracy, specificity, sensitivity, and F-measure are calculated. The adequacy of the proposed methodology is tested and verified on IEEE 4-bust test feeder and each ML algorithm is executed for 10 times due to different partitions and to prevent any overfitting result.

Copyright © 2020 Institute of Advanced Engineering and Science. All rights reserved.

1. INTRODUCTION

The widely used of nonlinear loads in power system have inject harmonics into the system and give rise to non-sinusoidal voltages and currents [1, 2]. The number and size of harmonic producing loads endlessly increase and lead to potential threat to the performance of power system. As a result, the harmonic distortion has increased losses and reduced the life expectancy of the equipments [3-5]. Therefore, it is important to define suitable methodologies to attribute the responsibilities for the deterioration of the power quality [6, 7]. There are numerous methods have been suggested by researchers to identify the location of harmonic sources based on different theoretical principles, features, benefits and drawbacks. Nevertheless, a high level of technical expertise and experience are required to properly diagnose the harmonic source type [8-10]. A studies on diagnosis method using stochastic diagnosis and power quality indexing been explained in [11, 12]. However, this method utilize Fourier and wavelet-transform signal processing techniques, whereas a lot of limitation of this signal processing techniques are explained in [13, 14]. Furthermore, a diagnosis study using short-time fourier transform (STFT) and S-transform based on rules-based classifier is introduced in [15-17]. Yet, high technical experience is required due to determine the rules-based classifier threshold parameter values [18]. Power quality analysis is a tricky task as the presence of complex linear as well as nonlinear patterns in the power quality event and old statistical method of auto-regressive integrated moving average (ARIMA) is introduced [19]. To improve the accuracy of ARIMA approach, the vector autoregressive (VAR)
model and artificial neural networks (ANN) are introduced. This is due to capture the linear interdependence and uncover the non-linear aspects, whereas made this method difficult and complex to apply. Many researchers used statistical learning theory (SLT) for pattern recognition. In [20], Kumar used neural network based classification algorithm for distinguishing the disturbance signals. However, this method is not suitable to construct frequency spectrum thus results in the loss of frequency components that has low energy components [21]. Fuzzy logic and adjusted probabilistic neural network are used in [22] to identify the location, level and type of harmonic sources however the total accuracy of this method is decrease whenever new type of load is introduced. Logistic regression is one the popular classification algorithms from the machine learning field. The logistic regression is widely used for classification compared to linear regression because its outcome on one sample is the probability that it is positive or negative and depends on a linear measure of sample [1, 23]. However this method cannot solve non-linear problems with logistic regression since its decision surface is linear and its high dependence on a proper data presentation [24]. This means that logistic regression is not a worthwhile tool unless all the important independent variables are available. To date, artificial intelligence such as machine learning has become one of the important techniques in the classification of the power quality system [25]. In the previous studies, it was found that the popular machine learning methods such as support vector machine (SVM), Naïve Bayes, neural network, and k-nearest neighbor were good for the classification and detection system, which can often provide satisfactory performance [26, 27]. Hence, machine learning can be useful in the classification of the power quality system. This paper proposes a high accuracy, fast estimation and costs efficient technique to identify and diagnose the type of multiple harmonic sources in the distribution system with single point measurement at the point of common coupling (PCC) by utilizing the machine learning techniques [28, 29]. The machine learning such as support vector machine (SVM) and Naïve Bayes are used to classify and diagnose the location and type of harmonic sources. SVM is one of the powerful machine learning methods, and it has widely used in other applications. Generally, SVM expands the concept of hyperplane separation to the data by expanding the feature vector in to a higher dimension [30]. Many types of kernel functions such as linear, Gaussian, and polynomial can be implemented in the SVM [31]. In this study, SVM with linear function is utilized, which capable to offer an excellent performance in current research [32]. NB is another popular and reliable machine learning method that has been widely used in pattern recognition studies. Generally, NB introduces the Bayes theorem to estimate the probability of data while all the features are assumed to be independent. Additionally, NB evaluates the most probable class when identifying the probability of feature vector [33]. At last, the best machine learning method for identifying and diagnosing harmonic sources is selected based on the performance measurement criteria.

2. RESEARCH METHOD

2.1. Machine learning methods

In this research, the machine learning methods are employed to identify the multiple harmonic sources. Prior to this work, two machine learning methods namely SVM and NB are utilized, mainly due to their promising performance in previous work.

3. PROPOSED METHOD

In most recent studies, suggest the execution of the proposed technique can be realized as depicts in Figures 1 and 2 using IEEE 4-bus test feeders, while the harmonic sources group into harmonic current source (HC) and harmonic voltage source (HV) type-load [34]. Four specific cases, which are case 1: N-N, case 2: N-H, case 3: H-H and case 4: H-N [16, 35, 36]. The main goal of this research is to diagnose and identify which case and type of the harmonic source in the power system.
where \( N \) is non-harmonic source which is resistor load, \( H \) is harmonic producing load and \( H \) whether \( H_C \) or \( H_V \), respectively.

Figure 3 shows the overview of identification and diagnosis of harmonic sources. Initially, the power quality signals which are current and voltage signals are measured at the PCC. In the next step, the S-transform is applied to transform the voltage and current signals into time-frequency representation (TFR). The signal parameters are then extracted from the TFR to form two feature sets: (1) current feature set, (2) voltage feature set. Furthermore, the features are normalized and then fed into the machine learning for the identification and diagnosis of harmonic sources. Besides, two popular and powerful machine learning algorithms, namely, support vector machine (SVM) and Naïve Bayes are applied to diagnose the NN, NH, HH, and HN cases for both \( H_C \) and \( H_V \).

3.1. Voltage and current feature sets

In this research, the voltage feature set contains five signal parameters that extracted from voltage and current signals of the PCC [37, 38]:

a) The average instantaneous RMS of voltage and current \( (V_{rms,ave} \text{ and } I_{rms,ave}) \)
b) The average instantaneous RMS fundamental of voltage and current \( (V_{1rms,ave} \text{ and } I_{1rms,ave}) \)
c) The average instantaneous total harmonic distortion of voltage and current \( (THD_{ave} \text{ and } THD_{iave}) \)
d) The average instantaneous total nonharmonic distortion of voltage and current \( (TnHD_{ave} \text{ and } TnHD_{iave}) \)
e) The average instantaneous total waveform distortion of voltage and current \( (TWD_{ave} \text{ and } TWD_{iave}) \)

3.2. Performance measurement of machine learning

Remark, several features are extracted from the power quality signals via S-transform and formed the two feature sets: (1) Voltage feature set, (2) Current feature set. After that, the features are normalized and then fed into the SVM and NB for the identification of multiple harmonic sources. Note that 10-fold cross-validation method is used for the performance evaluation. To measure the performance of proposed system, four evaluation metrics including accuracy, specificity, sensitivity, and F-measure are calculated [18, 31].

3.2.1. Accuracy

The accuracy is a metric that used to measure how accurate the proposed harmonic source diagnosis system can be. It is defined as,

\[
Accuracy = \frac{\text{No. of corrected diagnosed samples}}{\text{Total number of samples}}
\]

Figure 3. An overview of proposed method
3.2.2. Specificity
Specificity is another commonly used metric in diagnosis, and it can be calculated as,

\[
Specificity = \frac{TN}{TN+FP}
\]  

(2)

3.2.3. Sensitivity
Sensitivity is a commonly used metric in diagnosis, and it can be expressed as,

\[
Sensitivity = \frac{TP}{TP+FN}
\]  

(3)

3.2.4. F-measure
F-measure is an important metric that supports the accuracy, and it is used to characterize the performance of classifier. It can be defined as,

\[
F - measure = \frac{2TP}{2TP+FN+FP}
\]  

(4)

where TP is the true positive, TN is the true negative, FP is the false positive, and FN is the false negative, which can get from the confusion matrix.

4. RESULTS AND ANALYSIS
Table 1 presents the results of accuracy, sensitivity, specificity, and F-measure for the identification of the harmonic sources using SVM and NB for voltage feature set. In Table 1, it is clearly shows that the performance of the identification of harmonic sources was very low, which was below 40% accuracy. This means that the algorithms fail to estimate the harmonic source with problem. Based on the results obtained, voltage features were not good in describing the target concept, thus resulting in worst performance.

<table>
<thead>
<tr>
<th>Evaluation metrics</th>
<th>SVM</th>
<th>NB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.3906</td>
<td>0.3975</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.3036</td>
<td>0.3129</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.9005</td>
<td>0.9016</td>
</tr>
<tr>
<td>F-measure</td>
<td>0.2983</td>
<td>0.3072</td>
</tr>
</tbody>
</table>

Table 2 shows the results of accuracy, sensitivity, specificity, and F-measure for the identification of the harmonic sources using SVM and NB. As can be observed, by applying SVM and NB, different types of harmonic sources can be successfully identified. From Table 2, the overall performance of SVM was significantly better than NB in this work. The SVM gave an increment of 1.5% accuracy as compared to NB, which offered more accurate identification of harmonic sources. Moreover, the highest properties on sensitivity, specificity, and F-measure supported the results. As a result, it can infer that SVM is more appropriate to be applied in power quality system. As compared to voltage feature set, current feature set is more capable to identify the harmonic source effectively.

<table>
<thead>
<tr>
<th>Evaluation metrics</th>
<th>SVM</th>
<th>NB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.9750</td>
<td>0.9600</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.9786</td>
<td>0.9557</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.9959</td>
<td>0.9934</td>
</tr>
<tr>
<td>F-measure</td>
<td>0.9755</td>
<td>0.9561</td>
</tr>
</tbody>
</table>

Table 2 shows the results of accuracy, sensitivity, specificity, and F-measure for the identification of the harmonic sources using SVM and NB. As can be observed, by applying SVM and NB, different types of harmonic sources can be successfully identified. From Table 2, the overall performance of SVM was significantly better than NB in this work. The SVM gave an increment of 1.5% accuracy as compared to NB, which offered more accurate identification of harmonic sources. Moreover, the highest properties on sensitivity, specificity, and F-measure supported the results. As a result, it can infer that SVM is more appropriate to be applied in power quality system. As compared to voltage feature set, current feature set is more capable to identify the harmonic source effectively.

Figure 4 and Figure 5 illustrate the confusion matrix of SVM and NB for the identification of harmonic sources using voltage feature set. In these Figures, it shows that only N-N can be perfectly identified. Even though NB can work better than SVM, however, the performance is still far from perfect, which fail to identify the harmonic sources very well. Figure 6 and Figure 7 demonstrate the confusion matrix of SVM and NB for the identification of harmonic sources using current feature set. By applying
SVM, one can see that five types of sources (Hc-Hc, Hc-N, N-N, Hv-N, and N-Hv) were perfectly identified, in which a 100% class-wise accuracy was achieved. The harmonic source that most difficult to identify was N-Hc (89.29%). This is because SVM has misclassified the N-Hc as Hv-Hv, thus leading to worst performance.

![Confusion matrix of SVM using voltage feature set](image1)

![Confusion matrix of NB using voltage feature set](image2)

![Confusion matrix of SVM using current feature set](image3)

![Confusion matrix of NB using current feature set](image4)

5. CONCLUSION

In this research, the impact of current features and voltage features on the identification of multiple harmonic sources has been investigated. In addition, the performances of proposed feature sets are validated using popular machine learning methods, namely support vector machine (SVM) and Naïve Bayes. The experimental results indicate that the current features are more capable of enhancing the accuracy of harmonic source identification rather than using voltage features. Besides, our results prove the superiority of SVM in discriminating the seven types of harmonic sources, which contributes to the optimal performance in current work. All in all, the combination of current feature set and SVM is the most appropriate tools for the identification of harmonic source in power quality system.

ACKNOWLEDGEMENTS

This research is supported by the Advanced Digital Signal Processing Laboratory (ADSP Lab). Special thanks also to the Faculty of Electrical & Electronic Engineering Technology of Universiti Teknikal Malaysia Melaka (UTeM), Center for Robotics and Industrial Automation (CeRIA) of UTeM and Ministry of Higher Education Malaysia (MOHE). Their support is gratefully acknowledged.

Support-vector machine and Naïve Bayes based diagnostic analytic of harmonic ... (Mohd Hatta Jopri)
REFERENCES


BIOGRAPHIES OF AUTHORS

Mohd Hatta Jopri was born in Johor, Malaysia on 1978. He received his B.Sc fromUniversiti Teknologi Malaysia in 2000 and Msc. in Electrical Power Engineering from Rheinisch Westfälische Technische Hochschule Aachen (RWTH), Germany in 2011. Since 2005, he has been an academia staff in the Universiti Teknikal Malaysia Melaka (UTeM). His research interests are power electronics and drive, power quality, and DSP application.

Abdul Rahim Abdullah was born in Kedah, Malaysia on 1979. He received his B. Eng., Master Eng., Ph.D Degree from Universiti Teknologi Malaysia in 2001, 2004 and 2011 in Electrical Engineering and Digital Signal Processing, respectively. He is currently an Associate Professor with the Department of Electrical Engineering, Chief of Advanced Digital Signal Processing (ADSP) Lab and Center of Excellent (COE) Coordinator for Universiti Teknikal Malaysia Melaka (UTeM).

Jingwei Too received his B. Eng. in Mechatron Engineering from Universiti Teknikal Malaysia Melaka (UTeM) in 2017. He is currently pursuing his Ph.D. degree in the same university. His research of interest is bioengineering, signal processing, machine learning, deep learning, and feature selection.
Tole Sutikno is an Associate Professor in Electrical & Computer Engineering Department Universitas Ahmad Dahlan, Yogyakarta, Indonesia. He received B.Eng., M.Eng, and Ph.D in 1999, 2004 and 2016, respectively. His research interests include power electronics, motor drives, industrial electronics, Industrial application, FPGA application, intelligence control and industrial informatics.

Srete Nikolovski (M’95–SM’05) was born in Belgrade, Serbia, in 1954. He received the B.Sc. and M.Sc. degrees in electrical engineering from the Faculty of Electrical Engineering, University of Belgrade in 1978 and 1989, respectively, and the Ph.D. degree from the Faculty of Electrical and Computing Engineering, University of Zagreb, Croatia, in 1993. He is a Full Professor with the Department of Power Engineering, Faculty of Electrical Engineering, Computer Science and Information Technology, University of Osijek, Croatia. He has published 46 academic papers in journals and 164 papers in international conferences proceedings. His main interests are power system reliability, power system protection, integration of distributed generation in power networks, power system modeling and analysis, and arc-flash hazard risk analysis. He was a recipient of the IEEE PES Chapter Outstanding Engineering Award in 2014 for contributions in the field of power system reliability. He is an IEEE Senior Member of Reliability Society and PES Society.

Mustafa Manap was born in Kuala Lumpur, Malaysia on 1978. He received his B.Sc from Universiti Technologi Malaysia in 2000 and Msc. in Electrical Engineering from Universiti Teknikal Malaysia Melaka (UTeM) 2016. Since 2006, he has been an academia staff in the Universiti Teknikal Malaysia Melaka (UTeM). His research interests are power electronics and drive, instrumentation, and DSP application.