Non-Intrusive Load Monitoring Algorithm for PV Identification in the Residential Sector


Published in:

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Publisher rights
Copyright 2020 IEEE. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Open Access
This research has been made openly available by Queen's academics and its Open Research team. We would love to hear how access to this research benefits you. – Share your feedback with us: http://go.qub.ac.uk/oa-feedback
Non-Intrusive Load Monitoring Algorithm for PV Identification in the Residential Sector

Andres F. Moreno Jaramillo, David M. Laverty, Jesús Martínez del Rincón, Paul Brogan, D. John Morrow
School of Electronics, Electrical Engineering and Computer Science
Queen’s University Belfast
Belfast, United Kingdom
amorenojaramillo01@qub.ac.uk

Abstract—In response to the increasing penetration of distributed energy resources in the distribution network and the technical challenges this transition represents, this paper presents a novel approach for photovoltaic (PV) systems identification in the residential sector. Non-intrusive Load Monitoring (NILM) techniques have been focused mostly in identifying conventional loads on the customer side, thus more emphasis on distributed generation being integrated into the electrical grid is required to ensure system flexibility and most importantly stability of the electrical system. The proposed methodology combines basic statistics with the conventional machine learning Support Vector Machine, to identify PV load signatures from aggregated measurements in the residential sector using OpenPMU measurements. The main contributions of this paper are based on improving processing times of the conventional machine learning supervised algorithm and also providing important information for network operators based on simple techniques using electric current records from OpenPMU measurements.

Index Terms— Distributed Energy Resources, Non-Intrusive Load Monitoring, OpenPMU, Photovoltaic systems, Support Vector Machine.

I. INTRODUCTION

Consumers are increasingly acquiring photovoltaic (PV) roof top systems due to a variety of factors; concerns regarding climate change, advances in technology, kWp reduction and government incentives [1]. Electrical distribution networks are already facing issues regarding distributed generation, including constraints (infrastructure capacity) and more subtle issues including system losses, life time reduction of service transformers, reverse power flow, power quality reduction, and harmonic distortion [2].

In addition to PV systems, distribution networks are also experiencing changing operation needs due to the ever-increasing installed capacity of other ways of distributed generation, for instance battery energy storage schemes (BESS) and the integration of new low carbon loads as electric vehicles [3]. Collectively such technologies are known as Distributed Energy Resources (DERs).

Non-Intrusive Load Identification (NILM) has previously been demonstrated to be successful in identifying conventional loads. Here, a NILM method is developed to identify PV load profiles in the customer side. The main contribution of this paper is recognizing PV signature from aggregated measurements, providing valuable information for the distribution network operators (DNO) to enable them to respond to the new dynamics in the distribution network caused by integration of this DER. Implementing Support Vector Machine (SVM) as a supervised machine learning algorithm, OpenPMU [4] measurements, the NILM algorithm design on this work is based on basic statistical variables such as mean, median, variance and standard deviation as well as rms value of 50 samples moving windows. The dataset contains electrical current and voltage values from aggregated loads and a 3.5 kWp PV system from a residential installation in the United Kingdom.

The rest of this paper is organized as follows. Section II presents a brief overview of related NILM applications, followed by the description and mathematical representation of the models for the methodologies designed. Then, analysis and discussion of the main results obtained are provided in the fourth section of the document, including the used metrics. Finally, conclusion and future work are stated.

II. RELATED WORK

Load monitoring can be classified in intrusive and non-intrusive methodologies [5]. The first requires measurements of individual loads within an electrical installation, presenting high accuracy but also increased implementation cost as well as maintenance; whereas in NILM methods, aggregated measurements are used to identify individual load patterns also called electrical signatures, reducing the implementation costs and being useful for applications in power forecasting, demand side response and energy management systems [6].

One of the first NILM methodologies was introduced by George Hart in the late 80’s, based on the analysis of Active and Reactive Power (PQ) graphs to identify individual appliances from aggregated measurements. Although, simplicity and good performance of this methodology to
identify loads with different PQ characteristics, it is not reliable to classify for loads presenting similarities on this plane [7]. Therefore, many researchers have developed different methodologies considering conventional and non-traditional features observed in steady state or during transients. For instance, Gillis et al. presents in [8] a NILM algorithm based on discrete wavelet transformation for feature identification combined with a decision tree algorithm to identify four different loads during switching events, reporting an 96.65% accuracy. Additionally, Weisshaar et al. proposed in [9] a NILM using Frequency Invariant Transformation of Periodic Signals (FIT-PS), active and reactive power as inputs of conventional classifier algorithms such as k-Nearest Neighbor, SVM and Naïve Bayes. Finally, some of the methodologies has been developed to identify photovoltaic systems in the distribution network. Particularly, Dinesh et al. presented a methodology for load disaggregation and PV system identification plus generation prediction using at the residential sector [10]. Using an event based NILM method, it relied on Karhunen Loéve expansion (KLE) and sliding window to classify domestic loads. Implementing REDD, Tracebase and a dataset of 400 houses in California (USA), an overall accuracy above 80% was obtained.

III. PROPOSED METHOD

Non-Intrusive load monitoring algorithms are divided in three main parts: data acquisition, data processing and load classification as it is shown in Figure 1.

A. Data acquisition

An OpenPMU sensor was implemented as shown in Figure 2 to register the electrical current of a 3.5 kWp PV system, the aggregated loads and the voltage of a household in the United Kingdom. The sensor was installed from the 27th to 31st of May 2019, recording synchrophasors [11], thus providing magnitude and angle of each measured variable. There were implemented two channels to measure current of the PV system and the net demand of the house as well as the voltage of the residence. However, only PV and net load currents were analyzed due to this variable change depending on the characteristics of each system while the voltage remains almost constant all the time.

B. Data processing

At this stage, the dataset is sampled down from 50 Hz to 1 second resolution by taking magnitude of the synchrophasors and creating windows of 50 samples. The moving window is shifted half of the window size in order to cover any change in the load as it is demonstrated in Figure 4. In this way, major changes in the electrical current are recorded in the new matrix, including those variations located at the edge of each window.

![Figure 1. General diagram of NILM stages](image1)

![Figure 2. Data acquisition system configuration](image2)

![Figure 3. Demand, PV generation and Net load (27th to 31st of May 2019)](image3)

![Figure 4. Moving window process](image4)
In a second step, the mean, variance, standard deviation and \( \text{rms} \) values of each of the windows are calculated as follows [12].

**Mean.** Computing the sum of all aggregated load measurements from 1 to \( n \), with \( n \) equal to the window size and dividing by the total number of samples in the window the mean is defined by equation 1.

\[
\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}
\]  

(1)

**Variance.** Defined as the square deviation average from the mean, its mathematical equivalent is defined as indicated in equation 2.

\[
\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}
\]  

(2)

**Standard deviation.** Calculated as the square root of the variance, it is defined by equation 3.

\[
s = \sqrt{\sigma^2}
\]  

(3)

**\( \text{rms} \)** [13]. Defined as the effective value of the changing power during each window of time, the \( \text{rms} \) value can be described as shown in equation 4.

\[
\text{rms} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} x_i^2}
\]  

(4)

**C. Classification**

Presenting high accuracy, flexibility, robustness and flexibility SVM methods have been widely used in several NILM applications [16]. It is based on the creation of a function that better separates the features into classes, creating support vectors located at the maximum possible distance from a reference line, plane or hyperplane depending on the dimension of the input variables. Those vectors, can be described as presented in equation 6, where \( x_i \) represents the \( i \)th component of a particular feature, \( w \) is the weight vector and \( c \) the bias [17].

\[
w \cdot x_i + c = \pm 1
\]  

(6)

Translated into an optimization problem, the margin between the support vectors is required to be wide enough in order to correctly classify the \( x_i \) samples, which define the borders and general characteristics of the support vectors. The system is evaluated using labeled data, denoted as \( y_i \), which in this case contains the PV load profile information to calculate the efficacy of the classification algorithm.

**IV. ANALYSIS**

**A. Metrics**

This section presents the metrics used to evaluate the performance of the proposed NILM method using the experimental dataset, for both training and test stages.

Considering the numerous metrics available on the literature to estimate the efficacy of the identification algorithms, Precision (Pr), Accuracy (Acc), Recall (R) and F1-score (F1) have been chosen facilitating the comparison with other research studies outcomes [19] as well as providing a better understanding of the results which has set to be balanced.

The mentioned metrics can be expressed in terms of the amount of correct and incorrect predictions. More specifically, true positives (TP) and true negatives (TN) can be described as successful identification of the PV generating and PV out of service during a certain period respectively. In contrast, false positives (FP) are those predictions categorized as true while the PV is not delivering any power. Finally, false negatives (FN) are those attempts ranked as PV off while it is generating. For a better understanding, expressing the expected and predicted values combinations in terms of binary numbers, explanation of each class is provided in Table I.

**Table I. Metrics parameters**

<table>
<thead>
<tr>
<th>Class</th>
<th>Expected Value</th>
<th>Predicted Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TN</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>FN</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>FP</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>TP</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Then, each of the mentioned metrics can be mathematically derived from equations 7 to 10.

\[
P_r = \frac{TP}{TP + FP}
\]  

(7)

\[
Acc = \frac{TP + TN}{TP + FP + TN + FN}
\]  

(8)

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

(9)

\[
F1 = \frac{2 \times (P_r + R)}{P_r + R} = \frac{2 \times TP}{2 \times TP + FN + FP}
\]  

(10)

**B. Training and Test Sets**

In order to obtain a balanced dataset, the windows are labeled with a 1 for PV generating and a 0 when the PV off. To label a window including a transition of the PV from Off to On and vice versa, the current is analyzed during the second under study. For instance, if the current delivered by the PV is higher than 0.1 Amps for more than 50% of the second, the window is considered as PV generating, thus it is labeled with a 1.

Secondly, the total amount of windows containing PV generating and PV off are considered to create a balance training and testing sets. For this case, considering windows of 50 samples (1 second) values, around 62% of the data contained information of the PV generating while 38% were
windows with the PV out of service. Thus, only 38% of the PV generating windows were used to create both training and testing sets.

Finally, 75% of the dataset was implemented to train the machine learning algorithm and 25% to test it. Each of them was created using 50% of windows labeled as 0 and 50% labeled with a 1.

C. Method Evaluation

The proposed algorithm was evaluated using the experimental dataset created measuring electrical variables of a household in the United Kingdom with an OpenPMU. Synchro-phasors obtained at 50 Hz containing electrical current measurements have been selected to train and test the machine learning algorithm, implementing 5 days dataset with a grid connected PV system. The richness of this information is found on the diversity of the PV generation patterns during the observed period, producing a noon-constant bidirectional power flow and modifying the net load each day.

Two different experiments were carried out. The first one consists of using the windows as inputs of the system classifier and the second one was based on obtaining first the statistical variables of each window, reducing the size of the input from 50 samples to only 5. Both experiments where developed in python using an Intel Core i7-8700 CPU @ 3.20 GHz processor. The results obtained are presented in Table II.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Experiment 1</th>
<th>Experiment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
<td>75,842</td>
<td>76,834</td>
</tr>
<tr>
<td>TN</td>
<td>80,671</td>
<td>81,506</td>
</tr>
<tr>
<td>FP</td>
<td>3,255</td>
<td>2,415</td>
</tr>
<tr>
<td>FN</td>
<td>8,074</td>
<td>7,087</td>
</tr>
<tr>
<td>Precision</td>
<td>95.88 %</td>
<td>96.95 %</td>
</tr>
<tr>
<td>Accuracy</td>
<td>93.25 %</td>
<td>94.34 %</td>
</tr>
<tr>
<td>Recall</td>
<td>90.38 %</td>
<td>91.56 %</td>
</tr>
<tr>
<td>F1 Score</td>
<td>93.05 %</td>
<td>94.18 %</td>
</tr>
<tr>
<td>Processing Time</td>
<td>3,020 minutes</td>
<td>62 minutes</td>
</tr>
</tbody>
</table>

The closeness of the results in the metrics confirms the balance in the training and test sets. In both experiments, it can be seen how for the system it is slightly more difficult to differentiate cases of PV generating and classifying them as PV off, thus providing about 5 % FN and less than 2 % FP. This indicates the algorithm is not capable of completely recognizing low power generation from the aggregated measurements, but it is good at disaggregating the signals for a considerable PV power generation.

In addition, improving the results by 1% in all the metrics, the main contribution of using the statistical variables as inputs of the classifier rely on the considerable reduction in the processing time by a factor of 48.7. This is generated by reducing the size of each window from 50 to 5 samples comparing the first and second simulations.

C. Method Evaluation

The proposed algorithm was evaluated using the experimental dataset created measuring electrical variables of a household in the United Kingdom with an OpenPMU. Synchro-phasors obtained at 50 Hz containing electrical current measurements have been selected to train and test the machine learning algorithm, implementing 5 days dataset with a grid connected PV system. The richness of this information is found on the diversity of the PV generation patterns during the observed period, producing a noon-constant bidirectional power flow and modifying the net load each day.

Two different experiments were carried out. The first one consists of using the windows as inputs of the system classifier and the second one was based on obtaining first the statistical variables of each window, reducing the size of the input from 50 samples to only 5. Both experiments were developed in python using an Intel Core i7-8700 CPU @ 3.20 GHz processor. The results obtained are presented in Table II.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Experiment 1</th>
<th>Experiment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
<td>75,842</td>
<td>76,834</td>
</tr>
<tr>
<td>TN</td>
<td>80,671</td>
<td>81,506</td>
</tr>
<tr>
<td>FP</td>
<td>3,255</td>
<td>2,415</td>
</tr>
<tr>
<td>FN</td>
<td>8,074</td>
<td>7,087</td>
</tr>
<tr>
<td>Precision</td>
<td>95.88 %</td>
<td>96.95 %</td>
</tr>
<tr>
<td>Accuracy</td>
<td>93.25 %</td>
<td>94.34 %</td>
</tr>
<tr>
<td>Recall</td>
<td>90.38 %</td>
<td>91.56 %</td>
</tr>
<tr>
<td>F1 Score</td>
<td>93.05 %</td>
<td>94.18 %</td>
</tr>
<tr>
<td>Processing Time</td>
<td>3,020 minutes</td>
<td>62 minutes</td>
</tr>
</tbody>
</table>

The closeness of the results in the metrics confirms the balance in the training and test sets. In both experiments, it can be seen how for the system it is slightly more difficult to differentiate cases of PV generating and classifying them as PV off, thus providing about 5 % FN and less than 2 % FP. This indicates the algorithm is not capable of completely recognizing low power generation from the aggregated measurements, but it is good at disaggregating the signals for a considerable PV power generation.

In addition, improving the results by 1% in all the metrics, the main contribution of using the statistical variables as inputs of the classifier rely on the considerable reduction in the processing time by a factor of 48.7. This is generated by reducing the size of each window from 50 to 5 samples comparing the first and second simulations.

Finally, 75% of the dataset was implemented to train the machine learning algorithm and 25% to test it. Each of them was created using 50% of windows labeled as 0 and 50% labeled with a 1.

C. Method Evaluation

The proposed algorithm was evaluated using the experimental dataset created measuring electrical variables of a household in the United Kingdom with an OpenPMU. Synchro-phasors obtained at 50 Hz containing electrical current measurements have been selected to train and test the machine learning algorithm, implementing 5 days dataset with a grid connected PV system. The richness of this information is found on the diversity of the PV generation patterns during the observed period, producing a noon-constant bidirectional power flow and modifying the net load each day.

Two different experiments were carried out. The first one consists of using the windows as inputs of the system classifier and the second one was based on obtaining first the statistical variables of each window, reducing the size of the input from 50 samples to only 5. Both experiments were developed in python using an Intel Core i7-8700 CPU @ 3.20 GHz processor. The results obtained are presented in Table II.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Experiment 1</th>
<th>Experiment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
<td>75,842</td>
<td>76,834</td>
</tr>
<tr>
<td>TN</td>
<td>80,671</td>
<td>81,506</td>
</tr>
<tr>
<td>FP</td>
<td>3,255</td>
<td>2,415</td>
</tr>
<tr>
<td>FN</td>
<td>8,074</td>
<td>7,087</td>
</tr>
<tr>
<td>Precision</td>
<td>95.88 %</td>
<td>96.95 %</td>
</tr>
<tr>
<td>Accuracy</td>
<td>93.25 %</td>
<td>94.34 %</td>
</tr>
<tr>
<td>Recall</td>
<td>90.38 %</td>
<td>91.56 %</td>
</tr>
<tr>
<td>F1 Score</td>
<td>93.05 %</td>
<td>94.18 %</td>
</tr>
<tr>
<td>Processing Time</td>
<td>3,020 minutes</td>
<td>62 minutes</td>
</tr>
</tbody>
</table>

The closeness of the results in the metrics confirms the balance in the training and test sets. In both experiments, it can be seen how for the system it is slightly more difficult to differentiate cases of PV generating and classifying them as PV off, thus providing about 5 % FN and less than 2 % FP. This indicates the algorithm is not capable of completely recognizing low power generation from the aggregated measurements, but it is good at disaggregating the signals for a considerable PV power generation.

In addition, improving the results by 1% in all the metrics, the main contribution of using the statistical variables as inputs of the classifier rely on the considerable reduction in the processing time by a factor of 48.7. This is generated by reducing the size of each window from 50 to 5 samples comparing the first and second simulations.
REFERENCES


