FAULT DETECTION USING MACHINE LEARNING IN PV ARRAYS

Sunil Rao, Cihan Tepedelenlioglu (Co-PI), Devarajan Srinivasan* (Co-PI), Andreas Spanias (PI)
Arizona State University, *Poundra

Several faults occur in PV Arrays. These are caused by shading, soiling, inverter faults and manufacturing mismatches [1]. Data acquired during faults tends to luster in the feature space consisting of current, voltage and temperature measurements [4]. Certain aspects of our algorithmic and experimental research using this facility will focus on modeling faults interns of clusters and using machine learning algorithm to form and track these clusters.

PV arrays are reliable, but any fault which does occur is difficult to detect and repair. Studies of PV faults have shown a mean time to repair (MTTR) of between 3 and 19 days for conventional arrays with data collected only at the inverter. Clearly there is an opportunity to improve fault handling in PV arrays, using statistical signal processing methods, [2] on SMD data and which can lead to automated early detection and precise diagnosis of PV problems.

A. *Machine Learning in fault detection*

The use of machine learning in fault diagnosis can be formulated as a multiple hypothesis testing problem. Machine learning is useful for the detection and the identification of the type of the fault. For example, if one of the arrays receives less sunlight due to shading, machine learning could help identify the error in the shading conditions.

We have shown that fault detection can be performed using statistical outlier detection techniques [4]. However, performing diagnosis and localization of a fault is a much deeper problem. It requires data on array behavior under each fault condition. Moreover, PV arrays come in all shapes and sizes and may behave very differently from one another under similar fault conditions. A comprehensive PV fault dataset does not currently exist, since array operators are rarely involved in academic research and may wish to keep the performance of their systems proprietary. Gathering data from fault conditions is difficult, to obtain unless continuous monitoring is enabled. Finally, the overwhelming majority of arrays are fitted with I-V sensors only at the inverter, allowing minor faults which do not cause a large drop in output to persist undetected. Studies that attempt to quantify the likelihood and severity of different conditions were reported in [3]. On the other hand, extensive work has been done to characterize the behavior of normally operating modules and arrays [4].

A classification algorithm for fault detection must have the following properties. First it must accurately classify the PV array's condition. It must be adaptable to different array configurations without extensive data collection for each individual array. It must be able to recognize each fault class from a very small number of examples. It should take advantage of our prior knowledge of the electrical behavior of PV arrays (e.g. equal current within a string), rather than having to learn these relationships through the training data. It should be capable of reacting to the 'unknown unknowns' i.e. faults the system designers did not anticipate.

In light of these requirements, several machine learning approaches are worth examining. Semi-supervised learning could allow the generation of many realistic faults from a few measured examples. This would mitigate the problem of lopsided data, where very few examples of faults are available. Once generated, clustering approaches such as K-means or graph based semi supervised learning could be used to train a dataset]. We use the K-means algorithm as a starting point since there is an I-V level clustering observed in the data. Bayesian networks could allow researchers later to express known relationships and constraints on array behavior even if they are not displayed in training data.

B. *Using a simple K-means clustering algorithm*

The *k*-means algorithm (Figure 1) was chosen as an initial approach to machine learning-based fault detection. K-means clustering aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. Simulated fault data were obtained using the UW-Madison PV module performance module and a SPICE circuit simulation package. K-means algorithm was applied on these data.

The dataset was gathered under normal (well irradiated) conditions of temperature with high levels of current flowing through each panel. A cluster for each panel was formed based on the algorithm described below.

To simulate a shaded panel, one of the panels was assigned a lower irradiance value. The data for the same was obtained and trained with the k-means algorithm.

Description of the k-means Algorithm

Given a set of observations (x_1, x_2, …, x_n), where each observation is a d-dimensional real vector, k-means clustering aims to partition the n observations into $k(\leq n)$ sets S = {S_1, S_2, …, S_k} so as to minimize the inter-cluster sum of squares (ICSS) (sum of distance functions of each point in the cluster to the *K* centre:

The standard K-means equation is given, as follows: $\arg min \sum_{x \in Si} \sum_{i=1}^{k} \left|\left(x - \mu_i\right)\right|^2$ (1)

where μ_i is the mean of points in S_i.

We assign each observation to the cluster whose mean yields the least within-cluster sum of squares (WCSS). $S_i^{(t)} = \{x_p : \left\|x_p - m_i^{(t)}\right\|^2 \leq \left\|x_p - m_j^{(t)}\right\|^2 \forall j, 1 \leq j \leq k\}$ (2)

Each x_p is assigned exactly to one $S^{(t)}$. The centroid is updated in the next iteration as follows ,

$$m_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j \,. \qquad (3)$$

Where m_i is the updated mean.

This process is repeated for a pre-defined number of iterations.

For the PV simulation, we consider each μ_i to be the mean of the current measurement which is updated in every iteration of the set X. The set X represents the current reading at each panel over time. m_i represents the updated mean for the set S in each iteration. The set S is the number of clusters formed in each iteration *k*.

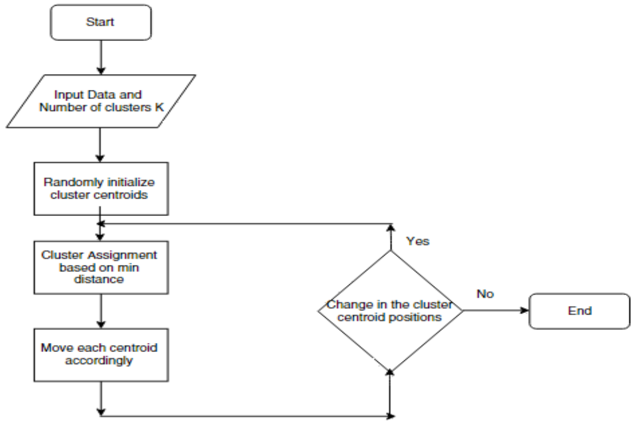

Figure 1: Flowchart demonstrating the operation of the K-Means algorithm.

I. PRELIMINARY RESULTS

Preliminary results indicate feature level separation between the data obtained from a faulty panel and data as obtained from a normal (working optimally) panel. This in the future could help in identifying the type of fault associated with each PV panel.

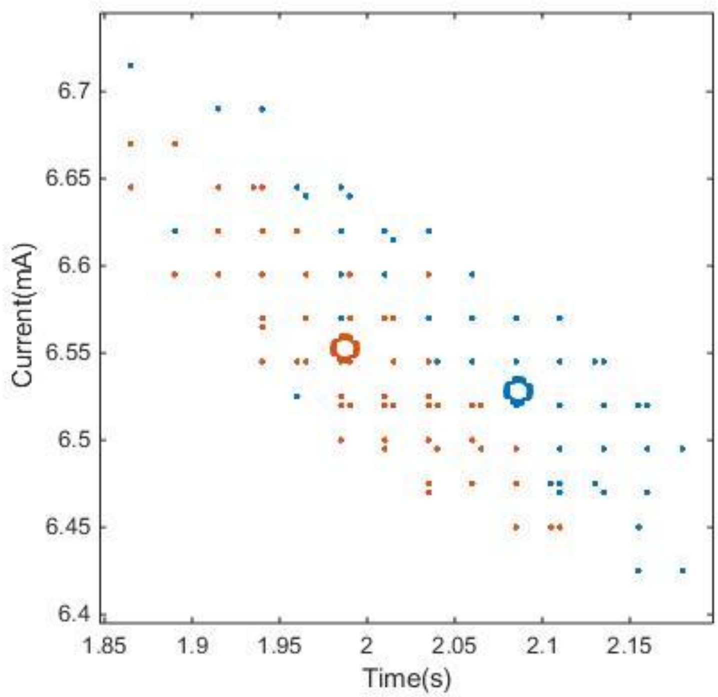

Figure 2: Preliminary results show separation between normally operating and faulty panels.

The results obtained at a preliminary stage are shown in Figure 2. Each data point in the feature space represents the measure of current by the PV panel. The faulty and non-faulty panels can be separated by means of a linear classifier. The centroids in the preliminary simulation shown in Figure 2 separate out well and tend to separate faulty and normal conditions.

II. REMARKS

Preliminary results on the use of machine learning for fault detection were reported in this report. The machine learning approach discussed here promises to detect faults occurring in PV arrays. The use of machine learning algorithms may be deployed as part of a robust monitoring system which improves the array efficiency with minimum human operator involvement.

References

[1] H. Braun,S.T.Buddha,V.Krishnan, A. Spanias, C. Tepedelenlioglu, T. Yeider,and T.Takehara, "Signal Processing For Fault Detection in Photovoltaic Arrays", *Proc. IEEE ICASSP 2012*, Kyoto, 2012.

[2] H.Braun, S.T.Buddha, V.Krishnan, A.Spanias, C.Tepedelenlioglu, T.Takehara, S.Takada, T.Yeider, and M.Banavar, Signal Processing for Solar Array Monitoring, Fault Detection, and Optimization, ser. Synthesis Lectures on Power Electronics, J.Hudgins, Ed, .Morgan & Claypool, vol.3, no.1, Sep.2012.

[3] A. B. Maish, C. Atcitty, S. Hester, D. Greenberg, D. Osborn, D. Collier, M. Brine, "*Photovoltaic system reliability", Photovoltaic Specialists Conference,* 1997.

[4] David L.King, "Photovoltaic Module and Array Performance Characterization Methods for All System Operating Conditions*", AIP conference proceedings,* 1997.

[5] Ye Zhao, Roy Ball, Jerry Mosesian, Jean-Franc¸ois de Palma, and Brad Lehman, "Graph-Based Semi-supervised Learning for Fault Detection and Classification in Solar Photovoltaic Arrays*", IEEE Transactions on Power Electronic*s, Vol. 30, No. 5, May 2015.