Abnormal Event Detection with High Resolution micro-PMU Data

Yuxun Zhou*, Reza Arghandeh†, Ioannis Konstantakopoulos*, Shayaan Abdullah*,
Alexandra von Meier*, Costas J. Spanos*
*Department of Electrical Engineering and Computer Sciences, UC Berkeley
{yxzhou, ioanniskon, sabdullah, vonmeier, spanos}@berkeley.edu
†Department of Electrical Engineering, Florida State University
arghandehr@gmail.com

***Abstract*—Power system has been incorporating increasing amount of unconventional generations and loads such as renewable resources, electric vehicles, and controllable loads. The induced short term and stochastic power flow requires high resolution monitoring technology and agile decision support techniques for system diagnosis and control. In this paper, we discuss the application of micro-phasor measurement unit (μPMU) for power distribution network monitoring, and study learning based data-driven methods for abnormal event detection. We first resolve the challenging problem of information representation for the multiple streams of high resolution μPMU data, by proposing a pooling-picking scheme. With that, a kernel Principle Component Analysis (kPCA) is adopted to build statistical models for nominal state and detect possible anomalies. To distinguish event types, we propose a novel discriminative method that only requires partial expert knowledge for training. Finally, our methods are tested on an actual distribution network with μPMUs, and the results justifies the effectiveness of the data driven event detection framework, as well as its potentials to serve as one of the core algorithms to ensure power system security and reliability.**

I. Introduction

A. Background

Historically, power distribution networks have not been equipped with sophisticated monitoring systems similar to what is implied in transmission networks. However, the growth of distributed renewable energy resources, electric vehicles and controllable loads introduces more short-term and unpredicted disturbances in the power flow [16]. This suggests a need for more accurate measurement devices with higher resolution. This paper specifically discusses high-precision synchrophasors, or micro-phasor measurement units (μPMUs) for high-fidelity measurement of voltage and current waveforms [1], which are designed to capture dynamic behavior of power distribution networks in order to support a range of diagnostic and control applications. All measurements are GPS time stamped to provide time-synchronized observability. μ-PMUs used in this research provide 120 samples per second for three-phase voltage and current magnitude and phase angle with a 0.05% Total Vector Error [15]. The accuracy and resolution available from this μ-PMU monitoring network enables operators to detect dynamic events that would otherwise be unobservable in distribution networks. Topology detection[2], phase labeling [17] and linear state estimation[13] are among applications of time synchronized μ-PMU data that are implemented so far.

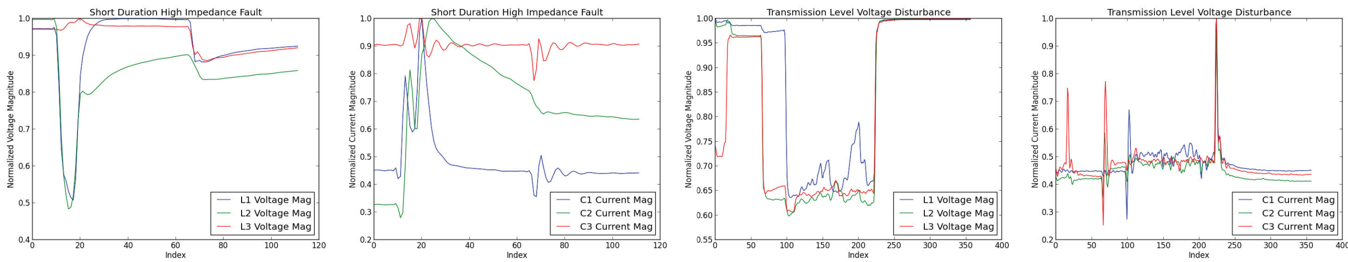
Events of interest in distribution networks are sinusoidal or non-sinusoidal transients in voltage and current waveforms that may be caused by faults, topology changes, load behavior and source dynamics. These events include, but are not limited to, voltage sags, voltage swells, fault currents, voltage oscillations, and frequency oscillations. For the sake of power systems reliability and stability, it is crucial to monitor the operating states in real time and detect anomalies quickly as to avert disturbances and disruptions[7]. Moreover, μ-PMU based monitoring system in distribution networks provides accurate and high fidelity data for a wide range of control strategies in different scales.

B. Learning Based Event Detection Methods

This paper proposes a novel framework for event detection using μPMU data streams. Within a larger scope, the paper aims to leverage advances in pattern recognition techniques and time-series data analysis for data-driven operation of distribution networks.

In literature, the model-based event detection has been very popular and successful in many applications as [10], [8]. However, it would be prone to overwhelming system randomness and dynamics in our context owing to the high time resolution and dimensionality. Therefore, the focus of the present work is to detect abnormal events based on μPMU measurements by considering only the empirical data *per se*. A wide variety of model-less detection and classification tools has been development recently developed in the machine learning/artificial intelligent community [6][9]. Most of model-less approaches require (1) useful information that is well presented and extracted as input “features”, and (2) training data that is well labeled with expert knowledge for supervised pattern recognition. However, these two prerequisites are becoming more challenging due to the nature of μ-PMU data that comes from the power distribution network with various components and diverse events. Moreover, identifying abnormal events in power distribution networks requires expert intervention which is not always available.

To resolve the first issue, we propose a pooling& picking scheme. Initially, all potential features are extracted with



(a) Voltage Magnitude: HI (b) Current Magnitude: HI (c) Voltage Magnitude: VD (d) Current Magnitude: VD

Fig. 1: Raw data visualization. 1 minutes PMU measurement for voltage and current that contains a short duration high impedance faults and transmission level voltage disturbanceis.

a sliding window and are pooled together as candidates. Secondly, a "Minimum Redundancy and Maximum Relevance (mRmR)" selection procedure is applied to pick useful information streams for each event. For the second issue, we propose two machine learning algorithms that perform unsupervised and semi-supervised detection depending on the availability of the labeled event data. The proposed procedure has following steps:

(1) Assumes only knowledge of the system's "stable state" or nominal condition is available. The kernel *Principal Component Analysis* (kPCA) is performed on the transformed Hilbert space to generate a tight non-linear description of the empirical distribution. Then the squared distance to the corresponding principal subspace is used as the measure of "abnormality" for a testing data point.

(2) Assuming that a small amount of data labeling can be achieved from the domain expertise, we propose a new variation of Support Vector Machine, the *Partially Hidden Structured SVM* (pSVM). The key idea is to incorporate information from all labeled, unlabeled and partially labeled data in a unified and large margin learning framework to reveal the hidden structures and capture relations among "stable state" and different "events".

The rest of the paper is organized as follows. In the next section, we discuss the pooling and picking procedure for information representation of the μPMU data. The event detection algorithms are described in Section III, and experimental results are given in Section IV.

II. Pooling and Picking: Information Representation and Feature Selection

The multi-stream high resolution PMU measurement data provides unprecedented observability of the system. Since all artificial intelligent methods are "garbage in, garbage out", the raw data that records values of a sensor measurement, must be properly processed for event related information before it is thrown into any machine learning algorithm. However, in this case facing milliseconds μPMU data that has not been explored before, we have limited prior knowledge on the effectiveness of different feature extraction methods. In this work, we propose a pooling& picking scheme: all plausible ways of feature extraction are firstly conducted on the μ PMU data, then the mRmR criterion is used to selected most informative ones for each type of events.

A. Pooling Candidate Features

Here we enumerate various information representation, i.e. feature extraction techniques that are related to the purpose of event detection. Notation-wise, the multi-stream time series μPMU data are written as $\{X_1, \cdots, X_T\}$. Each X_t is a $M \times C$ dimensional vector where M is the number of μPMUs and C is the number of channels of each μPMU. To get some intuition, two raw data sets containing 1 minutes 3 phase voltage and current measurements are shown in Figure 1 for a short duration high impedance (HI) faults and a transmission level voltage disturbance (VD).

Because the raw data is in millisecond's resolution and almost all practical events happens at a larger time scale, one can safely use a sliding window to extract useful information. The window size L should be chosen according to the time scale of the event of interest. For example, in order to detect certain transient event in 0.1s scale, one takes $L = 12$ and processes the data in each window. For ease of notation let $w_t^i \triangleq \{x_t^i, \cdots, x_{t+L}^i\}$ be the t^{th} window of stream i.

Now we consider diverse techniques to construct feature candidates. Intuitively, some events, such as voltage sag or voltage disturbance, could be revealed by investigating single streams (voltage magnitude or phase) fluctuations, while other events, such as high impedance fault and voltage oscillation, might be more obvious by analyzing the inter-behavior/dependence of multiple voltage and current streams. For the purpose of detecting different types of events, we include both single stream and inter-stream feature extract with a variety of metrics. To be specific, we consider

1) Single Stream Features Extraction:

- Classic statistics: including mean, variance, and range of voltage/current magnitude in each window. These features capture the average voltage/current values as well as their fluctuations in the time slot. The median is also included as it is a more "robust" metric of average value from a statistical viewpoint. To further characterize the variations of magnitude in each window, the distributional features, including entropy and histogram are calculated.
- First order difference: We compute $x_{t+1}^i - x_t^i$ for each stream and take the corresponding mean and variance in

TABLE I: Extracted Features Candidates

Single Stream	Statistics	$\text{mean}(w_t^i), \text{var}(w_t^i), \text{range}(w_t^i)$ $\text{median}(w_t^i), \text{entropy}(w_t^i), \text{hist}(w_t^i)$
	Difference	$u_t^i = \text{Diff}(x_t^i)$; Statistics
	Transformation	$\text{fft}(w_t^i), \text{wavelet}(w_t^i)$
Inter Stream	Deviation	$x^i - x^j \quad \forall i, \forall j \in \mathcal{N}(i)$
	Correlation	$corr(x^i, x^j) \quad \forall i, \forall j \in \mathcal{N}(i)$

each window. The intuition is that some transient events may exhibit significant "jumps" in voltage and current magnitude, which can be well captured by "spikes" in the first order difference. As for streams associated with phase information, the average difference is an indicator of voltage/current frequency and is also an important indicator of system stability.

- Transformation: Notice that many distribution side events, such as ON/OFF of reactive loads, usually lead to oscillations in both magnitude and phase measurement, we propose to use Fast Fourier Transform (FFT) to capture this frequency domain information. Also, Wavelet transformation is adopted to capture local fluctuations and abrupt changes, as is suggested in [14].

2) Inter-stream Features: For streams that corresponds to voltage/current of the same node, we compute

- Deviation: the difference between any two of the three phases, for both voltage and current. The resulted time sequences are processed as single streams in each window with classic statistics. In this way, we incorporate information for the events that exhibit phase imbalance.
- Correlation between any two of the three phases, for both voltage and current. The correlation constitutes a metric of dependence for these time series, and is also helpful in providing information related with inter-phase behavior.

A summary of feature extraction candidates are given in Table I. Note that the inter-stream features for different nodes (hence from different μPMUs) should be very interesting for sub-systems width event detection, for which one can include not only correlation as dependence metric, but also causal information [20] that pinpoints the propagation of the event. The task of identifying sub-system scale events and their influence on neighboring nodes is one of our future work. In this work, we will focus on data-driven method for local event detection.

B. Picking Informative Features

With the presented feature extraction procedure, a total number of $26C$ features have been pooled together. Obviously, some of them may be redundant as there are significant similarities among extracted features, for example, when the three phases are balanced, their single stream mean, variation, etc are almost the same. For another instance, the first order difference and wavelet transformation of one specific stream might have very similar pattern as they both reflect the sudden change of the same time series. From a machine learning point of view, adding redundant features does not help detection/classification, but instead would introduce extra learning noise and cause computational difficulties.

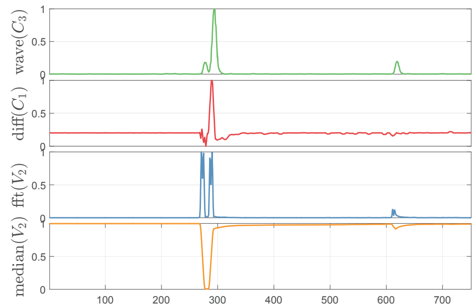


Fig. 2: Selected Features for HI

More importantly, for a particular event, or type of events, in practice only subsets of the calculated features are relevant, as it is mentioned earlier when those feature extraction techniques are proposed. After all, it is always beneficial to find out the "fingerprint" of each types of the event, not only for algorithmic concern, but also for system diagnosis purposes. In effect, the problem of feature selection, or variable subset selection, has long been an active research topic for statistics and machine learning. Existing methodologies could be divided into two categories and their combinations. The first type of methods define a feature contribution metric as an objective function, usually in terms of the dependency between candidate features and the target (data labels), thus reduce the problem into a combinatorial optimization. The second category views feature selection as trading off fitness and complexity of the learning model, based on which a series of regularized learning machines and optimal criteria have been proposed.

In this work, we adopt a combined method recently developed in [11], called Minimum-redundancy-maximum-relevance (mRMR). The procedure uses mutual information as the metric of goodness of a candidate feature set, and resolve the trade-off between relevancy and redundancy. To be specific, let $I(X;Y)$ be the mutual information between random variable X and Y, the first part of feature selection objective is to maximize the average dependence of selected feature set S on the target label c, i.e.

$$\max_{S \in \mathcal{X}} D(S) \triangleq \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c) \tag{1}$$

where we have denoted S as the set variable for a collection of features, and $\mathcal{X} = \{x^1, \cdot, x^d\}$ as the set of all candidate features. Considering that features selected only according to Max-Relevance criterion could have rich redundancy, a second objective, a penalty on average first order redundancy is introduced

$$\min_{S \in \mathcal{X}} R(S) \triangleq \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i, x_j) \tag{2}$$

Combining the above two consideration yields the mRmR feature selection objective

$$\max_{S \in \mathcal{X}} \{D(S) - R(S)\} \tag{3}$$

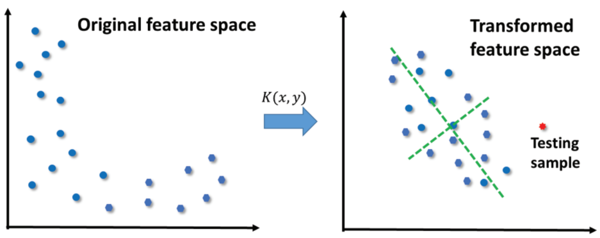


Fig. 3: Illustration of kPCA

which is approximately solved with a greedy heuristic: suppose we already have S_{m-1}, the feature set with $m-1$ features, then the next feature is found by optimizing the following one variable problem:

$$\max_{x_j \in \mathcal{X} \setminus S_{m-1}} \left[I(x_j; c) - \frac{1}{m-1} \sum_{x_i \in S_{m-1}} I(x_j; x_i) \right] \quad (4)$$

The mRmR procedure is implemented in C++, with a multilayer discretization technique for mutual information estimation [3]. For each events, we perform the selection method to choose 20 most informative features. The top 4 selected features for HI are shown in Figure 2.

III. Event Detection Algorithms

Depending on the availability of expert knowledge, we consider two goals of event detection with μPMU, and for each task we propose a corresponding algorithm to achieve data-driven decision making. The first task is to differentiate stable state and abnormal state, which is essentially a binary detection problem. In the second task, a more challenge problem of distinguishing event types is considered.

A. Abnormal Detection based on kPCA

The binary detection case has been extensively studied in literature and is referred to as abnormal (or novelty) detection. Indeed, many classic machine learning tools, such as Principle Component Analysis (PCA), Partial Least Squares (PLS), Independent Component Analysis (IDA), and Fisher Discriminant Analysis (FDA) have been widely applied in various fields. Readers are referred to [12] and the references therein for a comprehensive survey. In this work, we adopt kernel PCA [5] for the binary detection problem. Similar to PCA and IDA, the main idea is still to generate a simplified model (hence noise is removed) for the distribution support of the stable state, however with kernel method the computation is performed in a transformed space for non linear distributions. An illustration of the kernel mapping (feature transformation) and principle component models in the transformed space is shown in Figure 3.

Mathematically, we first map the data non-linearly into a higher dimensional space $\mathcal{F}$ with $\Phi \in \mathcal{H}$, yielding $\boldsymbol{x}_j \rightarrow \Phi(\boldsymbol{x}_j)$. After centering in the transformed space with $\widetilde{\Phi}(\boldsymbol{x}_j) = \Phi(\boldsymbol{x}_j) - \frac{1}{N}\sum_{j=1}^{N}\Phi(\boldsymbol{x}_j)$, the kPCA solves the following eigenvalue problem:

$$\lambda \left(\widetilde{\Phi}(\boldsymbol{x}_k) \cdot \boldsymbol{V} \right) = \left(\widetilde{\Phi}(\boldsymbol{x}_k) \cdot \widetilde{\boldsymbol{C}} \boldsymbol{V} \right) \quad \forall k \quad (5)$$

where $\widetilde{\boldsymbol{C}} = \frac{1}{N}\sum_{j=1}^{N} \widetilde{\Phi}(\boldsymbol{x}_j)\widetilde{\Phi}(\boldsymbol{x}_j)^T$ is the covariance matrix in the transformed space. Since all solutions $\boldsymbol{V}$ lie in the span of $\widetilde{\Phi}(\boldsymbol{x}_1), \cdots, \widetilde{\Phi}(\boldsymbol{x}_N)$, with the kernel trick we arrives at solving

$$N\lambda\boldsymbol{\alpha} = \boldsymbol{K}\boldsymbol{\alpha} \quad (6)$$

where $\boldsymbol{K}_{i,j} \triangleq \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) \triangleq \widetilde{\Phi}(\boldsymbol{x}_i) \cdot \widetilde{\Phi}(\boldsymbol{x}_j)$. Based on the solution and the relation

$$\boldsymbol{V}^l = \sum_{i=1}^{N} \alpha_i^l \widetilde{\Phi}(\boldsymbol{x}_i) \quad (7)$$

one can compute the reconstruction error of a testing data sample $\boldsymbol{z}$ as

$$\begin{aligned} E\left(\widetilde{\Phi}(\boldsymbol{z})\right) &= \left(\widetilde{\Phi}(\boldsymbol{z}) \cdot \widetilde{\Phi}(\boldsymbol{z})\right) - \left(\boldsymbol{V}^Q\widetilde{\Phi}(\boldsymbol{z}) \cdot \boldsymbol{V}^Q\widetilde{\Phi}(\boldsymbol{z})\right) \quad (8) \\ &= \kappa(\boldsymbol{z}, \boldsymbol{z}) - \frac{2}{N}\sum_{i=1}^{N}\kappa(\boldsymbol{z}, \boldsymbol{x}_i) + \frac{1}{n^2}\sum_{i,j=1}^{N}\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) - \sum_{l=1}^{Q} p_l(\boldsymbol{z})^2 \end{aligned}$$

with the projection of $\widetilde{\Phi}(\boldsymbol{z})$

$$\begin{aligned} p_l(\boldsymbol{z}) = \sum_{i=1}^{N} \alpha_i^l \Bigg[&\kappa(\boldsymbol{z}, \boldsymbol{x}_i) - \frac{1}{N}\sum_{r=1}^{N}\kappa(\boldsymbol{x}_i, \boldsymbol{x}_r) \\ &- \frac{1}{N}\sum_{r=1}^{N}\kappa(\boldsymbol{z}, \boldsymbol{x}_r) + \frac{1}{N^2}\sum_{r,s=1}^{N}\kappa(\boldsymbol{x}_r, \boldsymbol{x}_s) \Bigg] \end{aligned} \quad (9)$$

The reconstruction error is used as abnormality measure, as it represents the "deviation" of $\widetilde{\Phi}(\boldsymbol{z})$ to the top Q principle components of the normal case in $\mathcal{F}$. And a simple user specified threshold could be used as the decision rule.

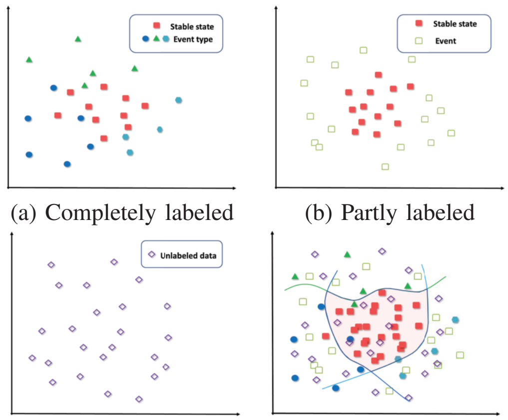
B. Semi-supervised Event Detection

In order to distinguish not only stable state from events but also different events types, more advanced data-driven machinery for multi-class classification is required. In the traditional supervised learning framework, data samples are firstly collected and labeled with expert knowledge for detailed system state (stable or event type), then a model is trained to describe the feature/label relationship. However, in the current μPMU based event detection application, expert knowledge for distribution system diagnosis is limited and scant. Hence, the availability of fully labeled data set is quite limited, which may cause insufficient learning and eventually lead to degraded detection performance. To alleviate this problem, we propose to incorporate information from partially labeled data, which does not require close scrutiny, as well as information from unlabeled data which is available in large quantity simply by collecting μPMU measurement of the system. To formalize the idea, let's define the following three data formats:

1 Completely labeled data samples, denoted as $\{\boldsymbol{x}_i, y_i, z_i\}$, where i is the sample index, $\boldsymbol{x}_i$ is the extracted features, y_i is the label indicating "stable" ($y = +1$) or an "event" ($y = -1$) with its type indicator $z_i \in \{1, \cdots, K\}$.

2 Partially labeled data samples, denoted as $\{\boldsymbol{x}_i, y_i, \cdot\}$, where y_i is still the indicator for "event", but no information about the event type is available.

3 Unlabeled data samples, denoted as $\{\boldsymbol{x}_i, \cdot, \cdot\}$, where only features $\boldsymbol{x}_i$ is accessible.



(a) Completely labeled (b) Partly labeled

(c) Unlabeled data (d) Porposed pSVM

Fig. 4: Different data format and intuition for pSVM

An illustration of these different situations is given in Figure 4. Intuitively, the partially labeled samples should be helpful: at least it provides discriminating information for binary decision making. The role of unlabeled data might be ambiguous at first glance since it does not carry any expert knowledge. However, it does contain distribution information in term of "clusters", as can be seen in Figure 4(c). To combine all three information sources and make best use of partial knowledge for event type identification, in this work we propose a new variation of the consensus learning framework [19], called the partial SVM (pSVM). Two new techniques are introduced to incorporate the effect of partially labeled data and unlabeled data.

For partially labeled samples, we see from Figure 1(a) and 1(b) that they could be viewed as data with "hidden event type". To compensate for this implicit information, we propose to describe the stable state by the intersection of the acceptance region of multiple base decision rules, while the events class by the union of their complements, as is illustrated in Figure 1(d). Mathematically, we write a composed classifier $g(\cdot)$ in terms of multiple base classifiers $f_k(\cdot)$ as follows:

$$g(\boldsymbol{w}, \boldsymbol{x}) > 0 \Leftrightarrow \min\{f_1(\boldsymbol{w}, \boldsymbol{x}), \cdots, f_K(\boldsymbol{w}, \boldsymbol{x})\} > 0$$

where K is the number of all possible event types. The multiple base classifiers performs implicit "clustering" for the event class, so as to capture hidden subgroups. Moreover, the classifier construction inherently emphasizes the sensitivity to event class as event is detected with any one of the base classifier, and it maintains the specificity to stable state class, as all base classifiers have to "agree" for a positive prediction. Again to consider non-linearity, a feature mapping $\phi : \mathbb{R}^d \to \mathcal{H}$ is applied. In the new Hilbert space a hyperplane classifier is written as $f(\boldsymbol{w}, \boldsymbol{x}) = \langle \boldsymbol{w}, \phi(\boldsymbol{x}) \rangle_{\mathcal{H}} + b \triangleq \boldsymbol{w} \cdot \phi(\boldsymbol{x})$ for short hand notation. Then the proposed classifier is

$$g(\boldsymbol{w}, \boldsymbol{x}) = \min_k \{\boldsymbol{w} \cdot \phi_k(\boldsymbol{x})\}.$$

and the hinge loss for a partially labeled data sample $\{\boldsymbol{x}_i, y_i, \cdot\}$ is just $[1 - y_i \min_k \{\boldsymbol{w} \cdot \phi_k(\boldsymbol{x_i})\}]_+$.

To include information provided by unlabeled data, the idea is to use a tentative labeling strategy $\widehat{y}_i = \operatorname{sign}(\min_k \{\boldsymbol{w} \cdot \phi_k(\boldsymbol{x}_i)\})$, then the corresponding hinge loss has the form

$$\left[1 - \widehat{y}_i \min_k \{\boldsymbol{w} \cdot \phi_k(\boldsymbol{x_i})\}\right]_+ = \left[1 - \left|\min_k \{\boldsymbol{w} \cdot \phi_k(\boldsymbol{x}_i)\}\right|\right]_+$$

Putting things together we propose the following regularized hinge loss minimization for event detection that incorporates all explicit and partial expert knowledge:

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & \frac{1}{2}||\boldsymbol{w}||^2_{\mathcal{H}} + c_1 \sum_{i \in \mathcal{L}^+} \left[1 - \min_k \{\boldsymbol{w} \cdot \phi_k(\boldsymbol{x_i})\}\right]_+ \\ & + c_{21} \sum_{i \in \mathcal{L}_1^-} [1 + \boldsymbol{w} \cdot \phi_{z_i}(\boldsymbol{x}_i)]_+ \\ & + c_{22} \sum_{i \in \mathcal{L}_2^-} \left[1 + \min_k \{\boldsymbol{w} \cdot \phi_k(\boldsymbol{x}_i)\}\right]_+ \\ & + c_3 \sum_{i \in \mathcal{U}} \left[1 - \left|\min_k \{\boldsymbol{w} \cdot \phi_k(\boldsymbol{x}_i)\}\right|\right]_+ \end{aligned} \qquad \textbf{(OPT1)}$$

where we have denoted $\mathcal{L}^+$ as the index set of all data samples that has $y_i = +1$, including both completely and cursorily labeled samples, $\mathcal{L}_1^-$ as the index set of completely labeled samples with $y_i = -1$ and event type z_i (hence the hinge loss only involves the corresponding individual classifier f_{z_i}). The index set $\mathcal{L}_2^-$ contains partially labeled samples in the event class, and $\mathcal{U}$ is the index of all unlabeled data samples. The loss penalty hyper-parameters c_1-c_3 weight each loss term differently, and should be chosen by taking into account the imbalanced cost for each scenarios.

The first three terms in the learning objective are convex in decision variables, the last two terms, however are not. To solve this challenging optimization problem, we transform **OPT1** into a joint optimization problem:

$$\begin{aligned} \min_{\boldsymbol{\eta}, \boldsymbol{\zeta}} \min_{\boldsymbol{w}} \quad & \frac{1}{2}||\boldsymbol{w}||^2_{\mathcal{H}} + c_1 \sum_{i \in \mathcal{L}^+} \left[1 - \min_k \{\boldsymbol{w} \cdot \phi_k(\boldsymbol{x_i})\}\right]_+ \\ & + c_{21} \sum_{i \in \mathcal{L}^-} [1 + \boldsymbol{w} \cdot \phi_{z_i}(\boldsymbol{x}_i)]_+ \\ & + c_{22} \sum_{i \in \mathcal{L}_H^-} \sum_{k=1}^{K} \eta_{ik} [1 + \boldsymbol{w} \cdot \phi_k(\boldsymbol{x}_i)]_+ \\ & + c_3 \sum_{i \in \mathcal{U}} \sum_{k=1}^{K} \zeta_{ik} [1 + \boldsymbol{w} \cdot \phi_k(\boldsymbol{x}_i)]_+ \\ & + c_3 \sum_{i \in \mathcal{U}} \zeta_{i(K+1)} \max_j \{0, 1 - \boldsymbol{w} \cdot \phi_j(\boldsymbol{x}_i)\} \\ \text{subject to} \quad & \boldsymbol{\eta}_i \in \mathbb{S}^K, \ \forall i \in \mathcal{L}_H^-; \quad \boldsymbol{\zeta}_i \in \mathbb{S}^{K+1}, \ \forall i \in \mathcal{U} \end{aligned}$$

The interchange of two minimization is justified because the optimization is bi-convex (quadratic in $\boldsymbol{w}$ and linear in $\boldsymbol{\eta}_i, \boldsymbol{\zeta}_i$) and is strictly feasible in both sets of decision variables. Interestingly, we see that for partially labeled data, we need K

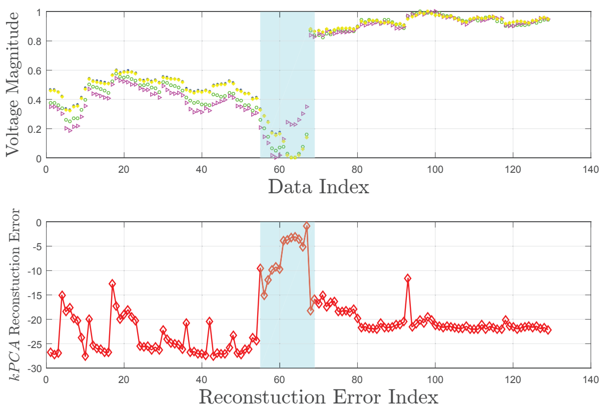


Fig. 5: Event Detection based on kPCA: voltage sag

indicators while for unlabeled samples, $K+1$ indicators are needed to distinguish 1 stable state plus K events. With the above bi-convex learning objective, many existing optimization heuristics can be readily used, such as group Alternating Optimization (AO) or Concave-Convex Procedure (CCCP), also more advanced tools can be adopted to approximate global optimum or reduce the computational time [4] [18].

IV. Evaluation

The authors are collaborating in the ”Micro-Synchrophasor for Power Distribution Networks” project [1] to install a number of μPMU devices at a number of distribution feeders. In this paper, actual data from some of the feeder installations are used to validate the proposed algorithm. The event detection by proposed two algorithms is performed for three phase voltage and current measurements at a substation on one of our installations.

A. Performance of kPCA

Due to the page limits, only one representative result for the binary detection with kPCA are presented here in Figure 5. In the experiment, we focus on the detection of potential voltage faults/fluctuation. Hence feature selection is performed for voltage related streams, and ten minitus data containing “stable state” samples are collected for building principle components in the transformed space. As for the calculation of reconstruction error, 20 principle components with largest eigenvalues are used. To test the distribution model we have built for stable state, we choose a new μPMU measurement sequence which contains a voltage sag from index 55 ($t = 5.5s$) to index 68 ($t = 6.8s$), shown as shaded area in the top plot of Figure 5.

The kPCA reconstruction errors for the testing sequence is shown on the bottom plot of Figure 5. We observe that the reconstruction errors for the abnormal region is significantly higher than the rest of the sequence. Hence with the algorithm one could easily identify the abnormality by specifying a simple threshold, which should be chosen with cross validation technique to balance the trade-off between sensitivity (which increases false alarm rate) and specificity (which increases miss detection rate). Moreover, it is also seen that the method could detect certain small scale voltage disturbances, such as the one around index 93 ($t = 9.3s$), as spikes can be found in the corresponding reconstruction errors.

TABLE II: Comparison of Detection Performance

Method	pSVM	Ada Boost.	Decision Tree	QDA
Accuracy	93.721	83.10	75.54	70.22

B. Performance of pSVM

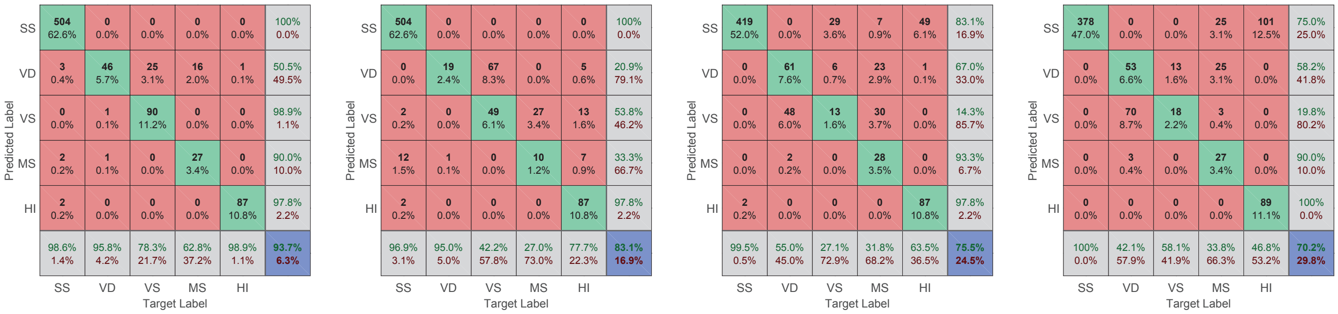
Next we evaluate the event detection performance of the proposed pSVM. In this experiment, we are interested in identify 4 different events, namely Voltage Disturbance (VD) and Voltage Sag (VS), Motor Start (MS), High Impedance fault (HI). The training set contains about 10 minutes μPMU records with detailed labels, and the testing data set also contains similar events, but is collect at a different time. For the training of pSVM, another 5 minutes partially labeled data and 10 minutes unlabeled data is included.

We compare the performance of pSVM with other popular multi-class classification methods, including Ada Boost, Decision Tree, and Quadratic Discriminate Analysis. Confusion matrices (contingency table) for all methods is shown in Figure 6. Each row of the table/figure represents the samples in predicted class while each column represents the samples in actual (true) class. The accuracy of each row/column is summarized in the left/bottom cell of the table, and the relative portion of the instances in percentage is shown underneath the main digit in each cell. We see that pSVM performs extremely well in distinguishing stable state and events, with 0% false alarm rate and only 1.4% miss alarm rate. Our method also works well in differentiating event types, especially VS, MS, and HI with an accuracy at least 90%. The only issue is that it tends to confuse VD with VS, which is somewhat expected as certain VD events are indeed quite similar to VS.

Comparing the results of pSVM to other methods, it is seen that in terms of overall accuracy, the proposed pSVM outperforms all the other methods, by at least 10.6% with respect to the runner-up Ada Boost, while a generic mixed Gaussian model used by QDA only yields 70.2% classification accuracy. When it comes to the performance of identifying event types, we see from the confusion matrices that pSVM is still the best algorithm: significant improvement is achieved for detecting VD and VS. and for MS and HI pSVM also has similar accuracy compared to the best of the other methods. The results justified the effectiveness of the proposed pSVM, as well as the idea of incorporating partial information for event detection.

V. Conclusion and Future Work

In summary, with the help of high resolution μPMU measurement, we have designed a pure data-driven framework for distribution network event detection in a refined granularity. The challenging problem of information representation of the raw data has been resolved with a pooling & picking procedure. Depending on the availability of expert knowledge, kPCA algorithm is adopted for binary decision making, and a novel learning method pSVM is developed to also distinguish event types by including information from both partially



(a) pSVM (b) Ada Boost (c) Decision Tree (d) QDA

Fig. 6: Confusion Matrix for comparison of methods

labeled and unlabeled data. The evaluation results justified the proposed methods.

For future work, we will implement kPCA and pSVM methods for a large volume of μPMUs from our site installations, as well as developing online learning pSVM. More recorded events will be used to train our algorithm and perform feature selection. More importantly, we will investigate the spatial-temporal characters of large scale events in power systems, and study data analytic tools for the corresponding detection problem.

Acknowledgment

The μPMUs project is sponsored in part by the U.S. Department of Energy ARPA-E program (DE-AR0000340). The authors would like to thank SinBerBEST. We also thank Kyle W Brady from UC Berkeley for his helps is gathering μPMUs data for event detection. Also, Ioannis C. Konstantakopoulos is a Scholar of the Alexander S. Onassis Public Benefit Foundation.

References

[1] Chapter 34 - every moment counts: Synchrophasors for distribution networks with variable resources. In L. E. Jones, editor, *Renewable Energy Integration*, pages 429 – 438. Academic Press, Boston, 2014.

[2] R. Arghandeh, M. Gahr, A. von Meier, G. Cavraro, M. Ruh, and G. Andersson. Topology detection in microgrids with micro-synchrophasors. In *Power Energy Society General Meeting, 2015 IEEE*, pages 1–5, July 2015.

[3] T. M. Cover and J. A. Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[4] C. A. Floudas. Nonlinear and mixed-integer optimization: Fundamentals and applications. Oxford University Press, 1995.

[5] H. Hoffmann. Kernel pca for novelty detection. *Pattern Recognition*, 40(3):863–874, 2007.

[6] M. Jordan and T. Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260, 2015.

[7] S. Meliopoulos, G. Cokkinides, R. Huang, E. Farantatos, S. Choi, and Y. Lee. Wide area dynamic monitoring and stability controls. In *Bulk Power System Dynamics and Control (iREP)-VIII (iREP), 2010 iREP Symposium*, pages 1–8. IEEE, 2010.

[8] P. A. R. A. Mohanty, S.R. A cumulative sum-based fault detector for power system relaying application. *IEEE Transactions on Power Delivery*, 2008.

[9] K. P. Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

[10] D. E. M. S. M. J. Napier, G. An automated fault analysis system for sp energy networks: Requirements, design and implementation. In *2009 IEEE Power and Energy Society General Meeting, PES*, pages 11–15, July 2009.

[11] H. Peng, F. Long, and C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(8):1226–1238, 2005.

[12] S. J. Qin. Survey on data-driven industrial process monitoring and diagnosis. *Annual Reviews in Control*, 36(2):220–234, 2012.

[13] L. Schenato, G. Barchi, D. Macii, R. Arghandeh, K. Poolla, and A. von Meier. Bayesian linear state estimation using smart meters and pmus measurements in distribution grids. In *Smart Grid Communications (SmartGridComm), 2014 IEEE International Conference on*, pages 572–577, Nov 2014.

[14] M. Sharifzadeh, F. Azmoodeh, and C. Shahabi. Change detection in time series data using wavelet footprints. In *Advances in Spatial and Temporal Databases*, pages 127–144. Springer, 2005.

[15] A. von Meier and R. Arghandeh. Diagnostic applications for micro-synchrophasor measurements, 2014.

[16] A. Von Meier, D. Culler, A. McEachern, and R. Arghandeh. Micro-synchrophasors for distribution systems. In *Innovative Smart Grid Technologies Conference (ISGT), 2014 IEEE PES*, pages 1–5, Feb 2014.

[17] M. H. Wen, R. Arghandeh, A. von Meiery, K. Poollay, and V. O. Li. Phase identification in distribution networks with micro-synchrophasors. In *Power Energy Society General Meeting, 2015 IEEE*, pages 1–5, July 2015.

[18] Y. Zhou, R. Arghandeh, I. Konstantakopoulos, S. Abdullah, and C. Spanos. Data-driven event detection with partial knowledge: A hidden structure semi-supervised learning method. *American Control Conference (ACC)*, 2016.

[19] Y. Zhou, N. Hu, and C. Spanos. Veto-consensus multiple kernel learning. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[20] Y. Zhou, Z. Kang, L. Zhang, and C. Spanos. Causal analysis for non-stationary time series in sensor-rich smart buildings. In *Automation Science and Engineering (CASE), 2013 IEEE International Conference on*, pages 593–598. IEEE, 2013.