Title
Abnormal event detection with high resolution micro-PMU data

Authors
Zhou, Yuxun
Arghandeh, Reza
Konstantakopoulos, Ioannis
et al.

Publication Date
2016-06-20

Peer reviewed

Distribution Network Event Detection with Ensembles of Bundle Classifiers

Yuxun Zhou*, Reza Arghandeh†, Ioannis C. 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 and Computer Engineering,CAPS, Florida State University
arghandehr@gmail.com
‡Ioannis C. Konstantakopoulos is a Scholar of the Alexander S. Onassis Public Benefit Foundation

***Abstract*—With the unprecedented growth of renewable resources, electric vehicles, and controllable loads, power system has been incorporating increasing amount of unconventional generations and loads. As a consequence, significant dynamic and stochastic power flow are introduced into distribution network, requiring high resolution monitoring technology and agile decision support techniques for system diagnosis and control. In this paper, we discuss the application of micro-synchrophasor measurement unit (μPMU) for power distribution network monitoring, and we propose a novel data-driven method, namely Ensembles of Bundle Classifier (EBC), for event detection. The main idea is: multiple classifiers are learned each with a short slot of μPMU measurement generated by a single event. Then their decisions are combined with a "winner-takes-all" scheme. This framework naturally resolves the challenging issue of heterogeneity in the high resolution μPMU data, and significantly outperforms classic data-driven event detection methods. In this paper, the proposed framework is tested on an actual distribution network with μPMUs, and is compared to other state-of-the-art methods. The result justifies the effectiveness of EBC as a promising tool to improve the security and reliability of distribution network.**

I. Introduction

A. Background

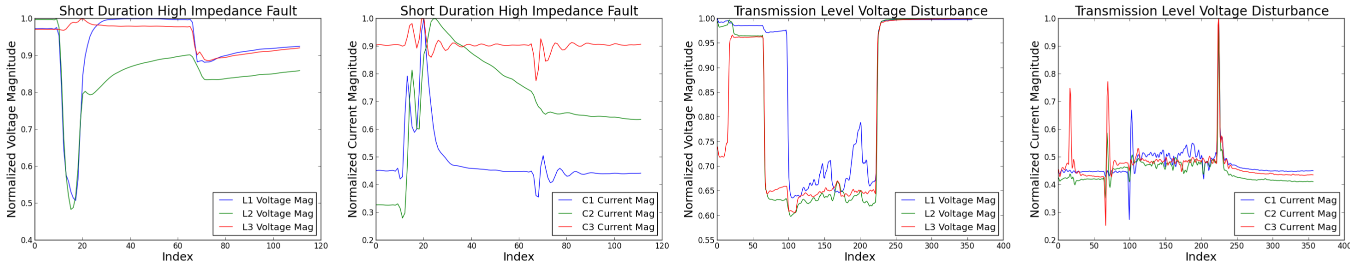
Historically, power distribution networks have not been equipped with sophisticated monitoring systems similar to what is implied in transmission networks. However, the growth of distributed renewable energy resources, electric vehicles and controllable loads introduces more short-term and unpredicted disturbances in the power flow [14]. 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 [13]. 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 [15] and linear state estimation[11] 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[6]. Moreover, μ-PMU based monitoring system in distribution networks provides accurate and high fidelity data for a wide range of control strategies.

B. Event Detection Methods

Event detection has been widely studied for many research disciplines and extensively used for various practical applications. It takes the measurement from system monitoring, and provides decision support for system control, diagnosis, etc. A large body of literature has been addressing the problem of event detection, and available works can be divided into two main categories and their combinations, i.e. model based method and model-less data-driven method.

The basic idea of model based approaches is to compare the system behavior, estimated by a dynamic model, to the expected behavior when the system is in certain state (normal or abnormal) [4]. Due to its simplicity and physical interpretability, the model-based methods have been very popular and successful in many applications, ranging from cyber attacks identification in power system [10], fault diagnosis for switching converters [3], and fault-detection of engine systems [12]. Nonetheless, this type of approaches rely heavily on correctness of the dynamical model of the system, as well as system analytic tools such as real time state estimators, parameter estimation, parity equations etc. Their limitations are obvious for the purpose of event detection with μPMU data: (1) dynamics of a voltage/current are hard to establish in millisecond time scale and (2) the overwhelming randomness in high time resolution and dimensionality makes any model unreliable.



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

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

Therefore, the focus of the present work is to detect abnormal events based on μPMU measurements by considering only the empirical data *per se*. The data-driven approaches use methods of machine learning/pattern recognition to conduct statistical inference or decision making on available system measurements [5]. Recently, many classic machine learning tools, such as kernel Principle Component Analysis (kPCA), Partial Least Squares (PLS), Independent Component Analysis (IDA), Support Vector Machine (SVM), and Fisher Discriminant Analysis (FDA) have been widely applied in various fields. Readers are referred to [9] and the references therein for a comprehensive survey.

However existing methods cannot be directly adopted for the current problem: In our view, the primary difficulty stems from the heterogeneity of the μ-PMU data that contains multiple events. Figure 1 shows two typical events, high impedance faults (HI) and voltage disturbanceis (VD) in power distribution network which constitute the target of our detection. We observe that the voltage/current measurements of these two events are quite different from each other. Nevertheless, classic methods essentially ignore this discrepancy and pool the measurement of various events together as positive class. This over-simplification is the major cause of degraded detection performance for classic methods.

To resolve the issue, we establish a novel data-driven event detection method for distribution network with μ-PMU measurement. In the first step of the proposed Ensemble of Bundle Classifier (EBC), a series of individual classification rules are learned, each with a short slot of event data as positive training sample and all stable state data as negative sample. Because one can choose a small enough slot length (bundle size) such that only one event occurs in that period, the learned classifiers are free from the aforementioned heterogeneity. As a second step, all individual classifiers are combined together: for a new sample, the classifier with smallest entropy (hence largest confidence) is taken for final prediction.

The rest of the paper is organized as follows. In the next section, we describe the EBC scheme and contrast it with classic method. In section III, we demonstrate the procedure of using probabilistic output and entropy criterion to combine multiple individual classifiers. Experimental results are given in Section IV, in which we also discuss the issue of information representation and compare EBC to other state-of-the-art methods.

II. Bundle Classifier for High Resolution data

The multi-stream high resolution μPMU measurement data provides unprecedented observability of the system. With milliseconds sampling interval, it can capture almost all variations in the network as real-life events usually happen at a larger time scale. On one hand, μPMU measurement enables the detection of various system events which would otherwise be indistinguishable with traditional power system monitoring technology. On the other hand, the high resolution and high dimensional data stream poses challenging issues for classic machine learning algorithms.

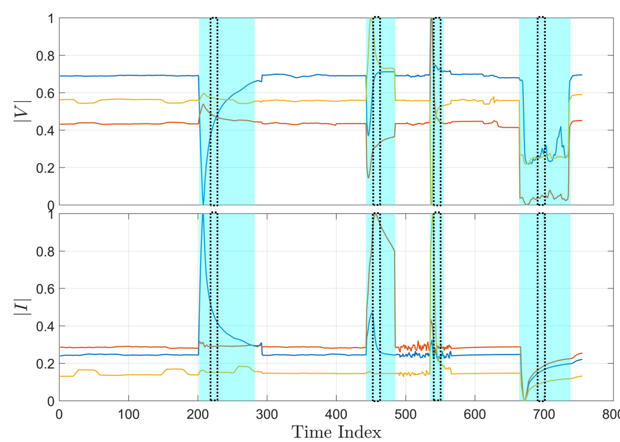
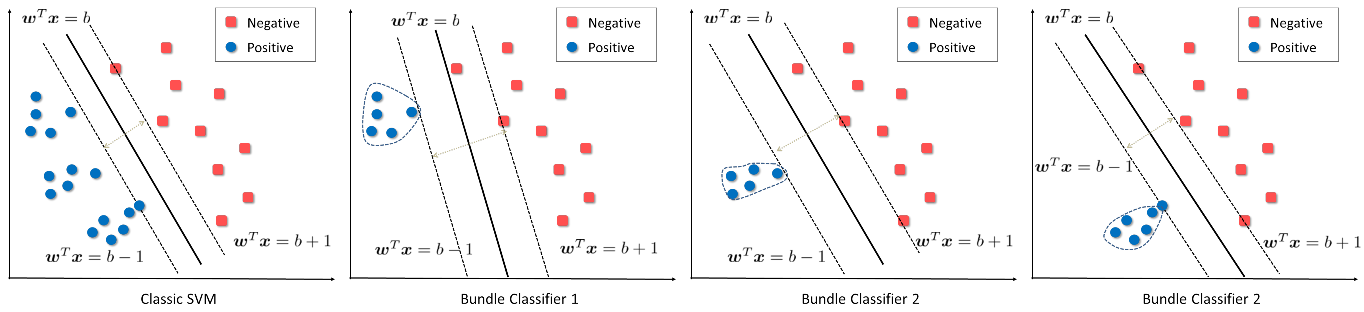


Fig. 2: A typical μPMU voltage and current measurements; 4 events in shaded cyan area; Sample bundles in dotted box

The key difficulty for event detection purpose, as we discussed ealier, is that the μPMU measurement associated with diverse events could have very different characteristics. To further elaborate this point, in Figure 2 we show a typical μPMU measurement for 3 phase voltage and current in 60 minutes[1]. The shaded cyan areas contains 4 common events, namely Motor Start (MS), High Impedance fault (HI), Voltage Sag (VS) and Voltage Disturbance (VD). The among-events discrepancies are obvious, moreover, event within a particular

[1] note that some stable state measurements between events are cut off for better illustration



(a) Classic SVM (b) Bundle Classifier No.1 (c) Bundle Classifier No.2 (d) Bundle Classifier No.3

Fig. 3: Classic machine learning tool vs. Ensembles of bundle classifiers

event the voltage/current responses are quite different from the beginning to the end of the event period. Classic machine learning methods, such as SVM, decision tree, or logistic regression, build a single classifier (decision rule), by pooling together measurement of events as positive training samples, and measurement of stable state as negative training samples. Figure 3a shows classic SVM as an example, in which the classifier is determined by maximizing the margin width of the separating hyper-plane. The major problem is that, the positive samples are actually from many subgroups, hence a single classification rule has to compromise among these subgroups for discrimination, which results in sub-optimal classifiers. In the SVM example, the algorithm is imposed to find a common weight $\boldsymbol{w}$ for features of all subgroups, leading to a small margin width of the separating hyper-plane which implies a mediocre classification performance.

The idea of bundle classifier is straightforward: We only take a small slot (bundle) of event data as positive training samples, as is illustrated in Figure 2 with dotted boxes. In these bundles, we expect the samples are approximately homogeneous and well concentrated. Then for each of the bundles, a classifier is built by contrasting the bundle of positive samples to all negative samples. An illustration is given in Figure 3b-3d. The bundle classifiers can be thought of as "expert" for each subgroups, which have optimized weighting coefficient and larger margin width.

In this work, we consider a recent variation of SVM, called 2ν SVM, to construct each bundle classifiers. The 2ν-SVM introduces an additional parameter $\gamma \in [0, 1]$ that decides the relative weight between the two classes errors. For the purpose of learning bundle classifier, we tune γ to compensate for the imbalanced training sample size in positive and negative class. The primal problem of the 2ν-SVM for $k^{th} \in \{1, \cdots, K\}$ bundle is given by

$$\begin{aligned} \min_{\boldsymbol{w}, b, \rho, \xi} \quad & \frac{1}{2}||\boldsymbol{w}||^2_{\mathcal{H}} - \nu\rho + \frac{\gamma}{n}\sum_{i \in \mathcal{I}^k_+} \xi_i + \frac{1-\gamma}{n}\sum_{i \in \mathcal{I}_-} \xi_i \\ \text{subject to} \quad & y_i(\kappa(\boldsymbol{w}, \boldsymbol{x}_i) + b) \geq \rho - \xi_i \quad \text{for} \quad i \in \mathcal{I} \\ & \xi_i \geq 0; \quad \rho \geq 0. \end{aligned} \tag{1}$$

where $\boldsymbol{x}_i \in \mathbb{R}^p$ are extracted features from the μPMU measurement, y_is are corresponding labels, and $\kappa(\cdot, \cdot)$ is the kernel function. $\mathcal{I}^k_+$ is the data index set for positive samples in the k^{th} bundle, and $\mathcal{I}_-$ is the index set of all available negative samples. Note that the above learning formulation (and its dual) is essentially a quadratic program, and can be solved efficiently with well established algorithms.

III. Combine Individual Classifiers

Having trained K bundle classifiers from data, the next step is to combine them together for final decision making. In the terminology of statistics and machine learning, this procedure is called "ensemble method". The key is to find reasonable "voting" scheme that utilizes information from all individual classifiers. Since in the current setting each bundle classifier is the "expert" for a specific case, we propose a "winner-takes-all" scheme, i.e., the final decision is made based on the most "confident" classifier.

To obtain a measure of "confidence" of a classifier, we first consider the probability output of the 2ν SVM, given by mapping the distance metric $f_i = \kappa(\boldsymbol{w}, \boldsymbol{x}_i) + b$ of sample i to a probability [8]

$$p_i(r, s) = \mathbb{P}(y = 1 | f_i) = \frac{1}{1 + \exp(r f_i + s)} \tag{2}$$

where r and s are model fitting parameters obtained by maximizing the log-likelihood on the training data set, i.e.,

$$\max_{r,s} \sum_{i \in \mathcal{I}} \{y_i \log p_i + (1 - y_i)\log(1 - p_i)\} \tag{3}$$

Performing the above calculation for all K bundle classifiers, we get K probability outputs for sample i, denoted as

$$\{p^1_i, \cdots, p^K_i\}$$

It is well known that entropy is a measure of inherent uncertain of a probability distribution. In other words, a more confident classifier should have smaller entropy on the predicted output. With this intuition, we compute the entropy

$$H(p^k_i) = -p^k_i \log_2 p^k_i - (1 - p^k_i)\log_2(1 - p^k_i) \tag{4}$$

for each of the probability output, and select the classifier with minimal entropy for prediction, i.e., $k^* = \text{argmin}_k H(p^k_i)$ and $\hat{y}_i = \text{sign}\{\kappa(\boldsymbol{w}_{k^*}, \boldsymbol{x}_i) + b_{k^*}\}$.

TABLE I: Extracted Features Candidates

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

IV. Experiment

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. Information Representation and Feature Selection

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, 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, firstly all plausible ways of feature extraction are conducted on the μ PMU data, then the mRmR criterion is used to selected most informative ones for each bundle.

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. 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. 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.

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 [16] 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. With the presented feature extraction procedure, a total number of 126 features have been pooled together. Obviously, some of them may be redundant as there are significant similarities among extracted features. For 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.

In this work, we adopt a feature selection method recently developed in [7], 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 feature selection objective is to maximize

$$\max_{S \in \mathcal{X}} \left\{ \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c) - \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i, x_j) \right\} \quad (5)$$

which is approximately solved with a greedy heuristic. 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 4.

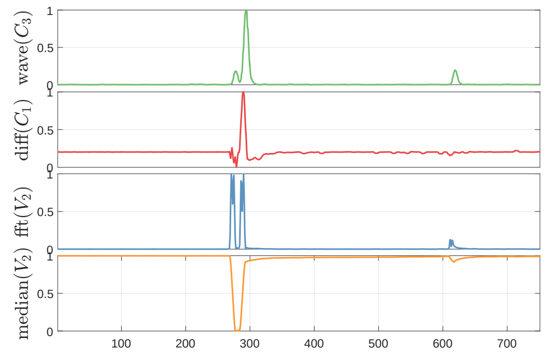


Fig. 4: Selected Features for HI

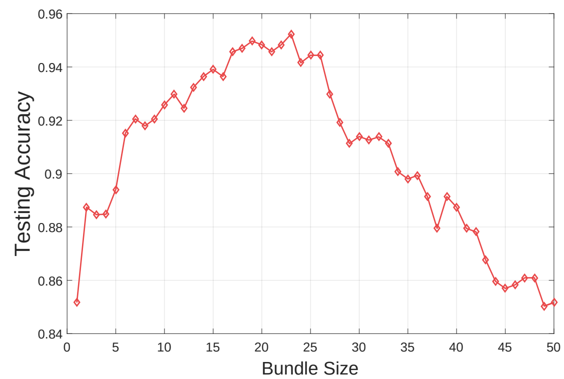


Fig. 5: Bundle Size vs. Testing Accuracy

B. Choosing the Optimal Bundle Size

Another important issue that is not addressed in previous discussion is the choice of bundle size. Intuitively, with small bundle size one can guarantee that the positive training samples are homogeneous (well clustered) for individual classifier, however, too small bundle size will lead to insufficient learning in the positive class. On the contrary, large bundle size can help increase training sample size of positive class, however one also face the increasing risk of including heterogeneous data in training.

To resolve this trade-off and find the optimal bundle size, we adopt a simple cross validation method. labeled data are

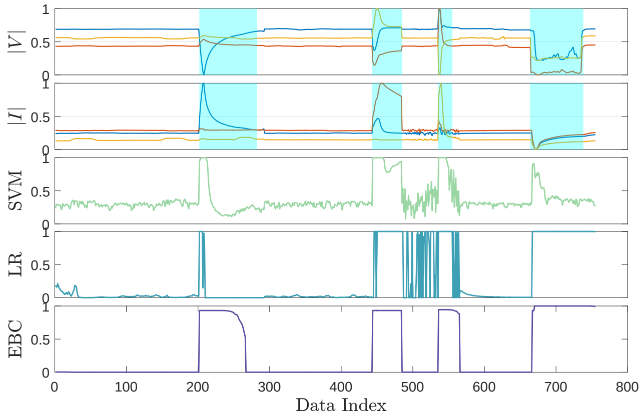


Fig. 6: Probabilistic Decision: SVM, LR and EBC.

divided into two subsets, a training set and a testing set. The EBC is learned on the training set with incremental bundle size, and for each scenario, the performance is evaluated on the testing data set. The testing accuracy as a function of bundle size is shown in Figure 5. We see that the result confirms our intuition: the testing accuracy initially increases with the bundle size, but eventually decreases as more heterogeneous data are included. The optimal bundle size for the μPMU data set is found to be 23.

C. Detection Performance of EBC

Finally, we demonstrate the detection performance of the proposed EBC and compare it with other popular classification methods, including Ada Boost, SVM, and Logistic Regression (LR). The predicted probability of event ($y = +1$) is illustrated in Figure 6, together with the testing data on the top two subplots. Comparing the prediction of each method to the ground truth (shaded cyan areas), we see that EBC performs extremely well in identifying various events and their boundaries, only with a few false alarms at the end of MS and VD event. Whereas SVM and LR generate considerable amount of miss detection in all four types of events. Moreover, the probability output of EBC is almost noiseless (hence is "confident"), while the prediction of other methods show fluctuations even for stable states.

The numerical results, including testing accuracy, false alarm rate (FAR) and miss detection rate (MDR) are shown in Table II. It is seen that in terms of overall accuracy, EBC outperforms all the other methods, by at least 9.54% with respect to the runner-up SVM, while LR only yields 80.66% detection accuracy. More importantly, EBC achieves 0% miss detection rate, indicating that all events are successfully detected. Its false alarm rate is also relatively low. The significant improvement compared to other methods justifies the idea of bundle classifiers, as well as the proposed minimum entropy voting scheme.

V. Conclusion and Future Work

In summary, we have designed a novel data-driven method for distribution network event detection in a refined granularity, with the help of high resolution μPMU measurement. We

TABLE II: Comparison of Detection Performance

Method	EBC	Ada Boost.	SVM	Logistic Regr.
Accuracy %	95.23	83.44	85.69	80.66
FAR %	9.54	17.10	6.73	9.72
MDR %	0.0	16.02	21.89	28.96

demonstrate with real experiment that the EBC framework can greatly improve the detection performance and is more stable in prediction. For future work, we will apply EBC to other detection problems. Besides, we will investigate the spatial-temporal characters of large scale events in power systems, and study data analytic tools for the corresponding event recognition problem.

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] X. Ding, J. Poon, I. Celanovic, and A. D. Dominguez-Garcia. Fault detection and isolation filters for three-phase ac-dc power electronics systems. *Circuits and Systems I: Regular Papers, IEEE Transactions on*, 60(4):1038–1051, 2013.

[4] R. Isermann. Model-based fault-detection and diagnosis–status and applications. *Annual Reviews in control*, 29(1):71–85, 2005.

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

[6] 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.

[7] 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.

[8] J. Platt et al. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, 10(3):61–74, 1999.

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

[10] B. M. Sanandaji, E. Bitar, K. Poolla, and T. L. Vincent. An abrupt change detection heuristic with applications to cyber data attacks on power systems. In *American Control Conference (ACC), 2014*, pages 5056–5061. IEEE, 2014.

[11] 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.

[12] A. Schwarte, F. Kimmidi, and R. Isermann. Model-based fault detection of a diesel engine with turbo charger-a case study. In *Fault Detection, Supervision and Safety of Technical Processes 2003 (SAFEPROCESS 2003): A Proceedings Volume from the 5th IFAC Symposium, Washington, DC, USA, 9-11 June 2003*, volume 1, page 293. Elsevier, 2004.

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

[14] 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.

[15] 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.

[16] 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.