1-1-2012

Synchrophasor Data Mining for Situational Awareness in Power Systems

Nischal Dahal

Synchrophasor data mining for situational awareness in power systems

By

Nischal Dahal

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Computer Engineering
in the Department of Electrical and Computer Engineering

Mississippi State, Mississippi

December 2012

Synchrophasor data mining for situational awareness in power systems

By

Nischal Dahal

Approved:

Roger L. King
Professor of Electrical and Computer Engineering
(Major Professor)

Ioana Banicescu
Professor of Computer Science and Engineering
(Committee Member)

Thomas H. Morris
Assistant Professor of Electrical and Computer Engineering
(Committee Member)

Yong Fu
Assistant Professor of Electrical and Computer Engineering
(Committee Member)

James E. Fowler
Professor of Electrical and Computer Engineering
(Graduate Coordinator)

Sarah A. Rajala
Dean of Bagley College of Engineering

Name: Nischal Dahal

Date of Degree: December 15, 2012

Institution: Mississippi State University

Major Field: Computer Engineering

Major Professor: Roger L. King

Title of Study: Synchrophasor data mining for situational awareness in power systems

Pages in Study: 118

Candidate for Degree of Doctor of Philosophy

Recently, there has been an increase in the deployment of Phasor Measurement Units (PMUs) which has enabled real time, wide area monitoring of power systems. PMUs can synchronously measure operating parameters across the grid at typically 30 samples per second, compared to 1 sample per 2-5 seconds of a conventional Supervisory Control And Data Acquisition (SCADA) system. Such an explosion of data in power systems has provided an opportunity to make electrical grids more reliable. Additionally, it has brought a challenge to extract information from the massive amount of data.

In this research, several data mining algorithms are used to extract information from synchrophasor data for improving situational awareness of power systems. The extracted information can be used for event detection, for reducing the dimension of data without losing information, and also to use it as heuristic to process future measurements.

The methods proposed in this research work can be broadly classified into two parts: a) stream mining and b) dimension reduction. Stream mining algorithms provide solution utilizing state-of-the-art data stream mining algorithms such as Hoeffding Trees (HT). HT algorithm builds a decision tree by scanning the incoming data stream only

once. The tree itself holds sufficient statistics in its leaves to grow the tree and also to make classification decisions of incoming data. Instead of using a large number of samples, which leads to a tree too large to accommodate in memory, the number of samples that are needed to split at each node is determined using Hoeffding bound (HB). HB keeps the size of the decision tree within bounds while also maintaining accuracies statistically competitive to traditional decision trees.

Dimension reduction algorithms reduce dimension of the synchrophasor data by extracting maximum information from a huge data set without losing information. In this dissertation, both online and offline dimension reduction algorithms have been studied. The online dimension reduction uses an unsupervised method using principal components of the time series data. The offline method optimizes unique mutual information between the state of the power system and synchrophasor measurements. It optimizes the criteria by reducing redundant information while maximizing relevant information.

DEDICATION

To my wife Pooja, Dad Narayan Dahal, Mom Bidhya Dahal and sister Namrata.

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Roger King for his continuous guidance and encouragement throughout this work. This work would not have been possible without his support and belief in me. I would like to acknowledge Dr. Ioana Banicescu, Dr. Thomas Morris and Dr. Yong Fu for their advice and support as members of my graduate committee. I would like to take this opportunity to show my gratitude towards Mr. Vahid Madani and the Pacific Gas and Energy (PG&E) Company for supporting this work.

I would like to acknowledge Dr. Surya Durbha and Mr. Uttam Adhikari for their help and support to finish this work in time.

I am very lucky to have Pooja as my loving wife, who has never stopped in believing me and supporting me in difficult times and rejoicing my successes. I would like to thank my sister Namrata and brother-in-law Bisesh for her love and support, which has always been encouragement for me. I would like show my respect to my parents for their unconditional love. I am really blessed to have their love and support throughout my life. I also would like to thank sister-in-law Pragya and brother-in-law Yam for making it easy to take difficult decisions.

Finally, I would like to take this opportunity to show my sincere gratitude towards all the teachers, professors, relatives and everyone who directly/indirectly supported my work, education and research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Background

The electric grid is considered by some to be the greatest engineering achievement of the 20th century [1]. However, even the 99.97% reliability of the present electrical system is not enough to prevent $150 billion in losses from outages and interruptions [2]. With the increasing dependence of human enterprises on electrical energy, grid reliability has emerged to be one of the most important factors in economic security. Unfailing reliability is a formidable challenge and is highly dependent upon the delicate balance of instantaneously matching generation to load. The facts that electrical energy cannot be stored in large scale, is dependent upon an indefinable real-time user demand profile, and is governed by laws of physics make it one of the most complicated systems to control. It is a formidable technological challenge to maintain a continuous supply of high quality energy from generation facilities to a consumer's appliances.

Very little has changed in the operation of a power system since it was invented in the 1880s [2]. Generation is still primarily fossil fueled with centralized plants, tariffs are still dominated by simplistic rate structures, and consumers still have little knowledge about the status of the grid [2]. With the enactment of the Energy Independence and Security Act of 2007 [3] there began a move toward making the future electric grid more efficient, environmental friendlier and more reliable by using modern technologies. The electricity industry is on the verge of a major paradigm shift (smart grid) in its operations by the joint efforts of utilities, industry, universities and the government.

The smart grid initiative envisions several goals for the future electric grid. First, the integration of renewable energy sources into the grid. Second, a two way exchange of information (e.g., real time tariffs) between utilities and consumers so that appliances can smartly shift their load to cheaper off-peak hours. Third, consumers not only consume energy, but also can produce and sell energy produced with photovoltaic (PV) cells mounted on their roofs or wind turbines on their property or even from an electric vehicle parked in their garage. All of these distributed sources of generation may help utilities to meet energy demand in peak hours [1]. The annual cost of meeting the demand of the highest 100 peak demand hours is about 10-20 percent of the entire electricity generation cost for the year. The measures to meet peak energy demand envisioned in the smart grid can save utilities billions of dollars in a single year in terms of transmission line congestion cost, savings in construction of new generation plants, etc [4]. The improvement of operation of the power system will have a profound impact on many areas of society; it has been viewed as having a similar impact as the interstate highway system had in the 1960s [5].

Utilities need to maintain balance between load and generation to maintain reliable operation of the power system. The integration of renewable energy sources, de-regularization, and two way communications between consumer and utilities make operation of the power system extremely complicated to handle. Currently, power system operations depend on preventive measures that would help the system sustain forecasted contingencies without losing synchronism of generators or voltage stability. The contingency-based model of protection may not be reliable in a grid where a significant portion of energy is supplied by variable energy sources; such as wind and/or solar. A real time awareness of the system is required so that corrective actions could be taken if any condition, which may lead to instability, is detected [4].

Motivation

Various control and protection schemes have been designed to control a wide range of dynamic phenomena of the power system. Some characteristics are of a very fast changing behavior and may last only a few milliseconds while others are slow moving characteristics [6]. In the case of fast moving phenomena, conventional control and protection mechanisms take over the decision making process without human intervention in the loop (reaction time is less than 100 milliseconds). For long term stability issues, operators usually have enough time to study historical data patterns, run simulations and consult their team before making any crucial decisions. But, there are times between these two extremes in which operators have to use their own judgment to act on certain conditions and this often happens when there is insufficient information available to support their decisions [7]. Figure 1 shows response timing requirements for different events in power systems.

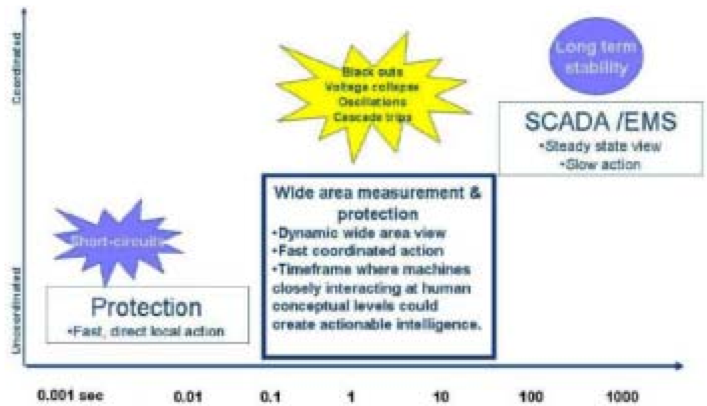

Figure 1 Timing requirement for different events [7]

Most of the control mechanisms are based on local measurements while most of the system's dynamic behavior depends upon the broader status of the power system (regional or system-wide) [6]. The increasing trend in the deployment of Phasor Measurement Units (PMUs) has enabled the power industry to have a wide area monitoring capability consistent with the scope of the dynamic behavior drivers. PMUs can synchronously measure operating parameters across the grid typically at 30 samples per second, compared to 1 sample per 2-5 seconds of a conventional Supervisory Control and Data Acquisition (SCADA) system. The availability of time synchronized high speed data from a suitably placed network of PMUs can capture the dynamic performance of the power system, which is not possible with conventional SCADA systems. A Wide Area Monitoring System (WAMS) acquires Global Positioning System (GPS) synchronized current, voltage and frequency measurements via optimally placed PMUs for maximum observability [8] (see Figure 2). The measured quantities include both magnitudes and phase angles, and are time-synchronized via GPS receivers with an accuracy of one microsecond. The time aligned measurements from the WAMS gives a snapshot of power system operations at any given time, so that both dynamic real time monitoring and post event analysis of the power system can accurately be done.

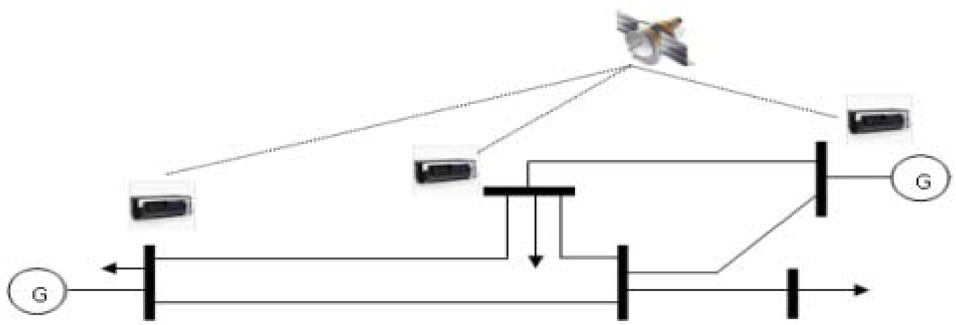

Figure 2 Wide area monitoring of power systems

In the past, Wide area monitoring parameters, such as phasor angles, could only be estimated after numerous iterations of power flow solutions, but they can now be directly measured with PMUs. The investigation of the August 14, 2003 blackout pointed out that the blackout could have been prevented if phasor data had been monitored. A number of clues surrounding the blackout were missed due to a lack of infrastructure to provide operators awareness of the severity of the situation. The phasor angle difference between Cleveland and Michigan, as shown in Figure 3, showed a significant divergence from normal. If this divergence had been detected in time, then proper actions may have prevented the entire blackout, or at least limited its spread to a smaller area [9].

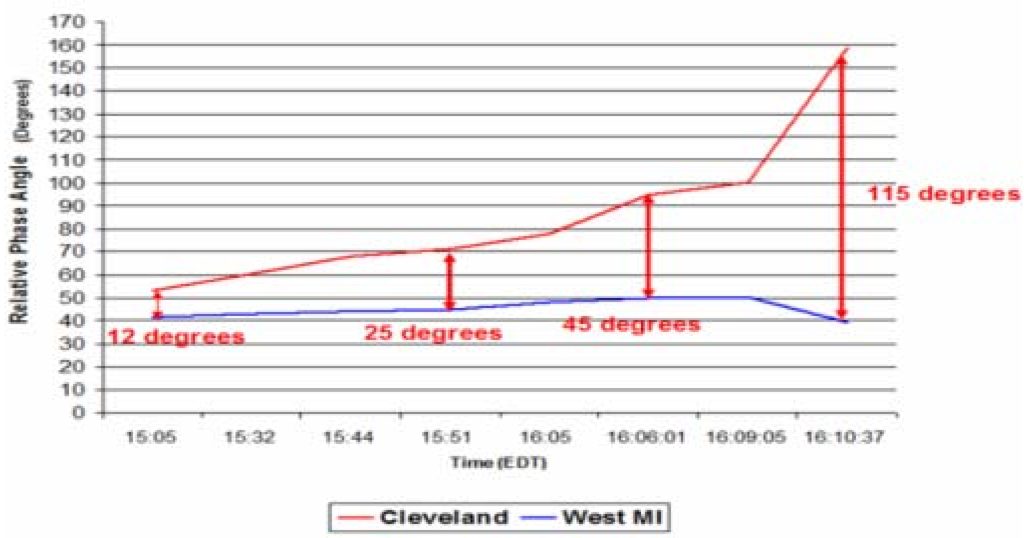

Figure 3 Angle separation between Cleveland and West Michigan on August 14 2003 [9]

PMUs have empowered industry with the capability of monitoring grid health parameters in real time. However, it is a challenge to extract information from a high speed data stream for situational awareness inside the control room. Typically, off-line mathematical calculations (e.g., power flow solution) are considered to be reliable for forecasting/predicting behavior of power system. However, the time required for mathematical calculations makes this approach infeasible for real time situational awareness applications. As an alternative to the accurate mathematical model, researchers have been studying different machine learning techniques that will help predict events on the grid within an actionable time frame.

The use of machine learning techniques is not a recent development in power system research rather it has been studied since 1960s [10]. Several machine learning techniques such as Support Vector Machines (SVM) [11-15], Artificial Neural Network (ANN) [16-21], and Decision Trees (DT) [22-26] have been used for several applications of static/dynamic security assessment and fault detection/classification. The accurate use of machine learning in predicting behavior of highly nonlinear, complex physical phenomena of large dynamic systems (such as power system) can have a significant contribution, particularly in the case of real time situational awareness applications where a compromise between accuracy and speed is needed.

Objectives

Phasor measurement units (PMUs) provide a continuous stream of time synchronized data about grid operating parameters. The synchronization of data enables operators to see a snapshot of the status of the power system in real time. It also helps in post event analysis to provide insights into exact cause of an event/outage which was not possible with traditional SCADA systems.

As synchrophasor data is becoming more available, power system researchers are experimenting with new ways of utilizing the continuous stream of data to enable more reliable, efficient and green decisions. A number of publications can be found (See Chapter 2) on using machine learning and pattern recognition methods to speed up event detection and classification in power systems. In this dissertation work, information extraction methods for synchrophasor data using different machine learning and pattern recognition methods are studied.

Traditional machine learning algorithms are designed to work on a small amount of data; predictive models are created with multiple scans of training data [27]. Models created using this approach can represent search space in available memory at an acceptable computational requirement for recall. The traditional algorithms for machine learning may not be appropriate for continuous streams of synchrophasor data because of the following reasons

- *Massive Data:* PMUs give a stream of data at 30 samples per second. Generally, the operation of the power system is considered in a period of 24 hours, which comprises of more than two million daily samples of data for each monitored parameter for each PMU deployed. The amount of data increases exponentially as more PMUs are deployed. The models created by traditional data mining algorithms increase in size and computational efficiency of recall degrades, negating the advantage of using machine learning algorithms instead of analytical methods.
- *Dynamic Behavior of Power System:* The power system is a very dynamic system. Operating Conditions (OC) of the power system change all the time. Even events such as opening of breakers change OC of power system. Traditional machine learning algorithms learn a model and make predictions. The learned knowledge is not easily updated without "forgetting" the previously learned knowledge and retraining the model. This approach would be impractical and very difficult to pursue in a dynamic system like the electric grid. An incremental learning method which can adapt to changing conditions without unlearning previously acquired knowledge is required.

Stream mining algorithm is a new approach to real time data mining of continuous streams of data similar to the synchrophasor data stream. This method creates a model of the system being emulated that can fit into memory and be recalled within acceptable latency criteria regardless of the number of training examples. In addition, this method also sought an incremental learning method to comprise the dynamicity of behavior of the system by updating the learned model frequently without forgetting previously learned knowledge.

One of the major objectives of this work is to prove the effectiveness of data stream mining algorithms for event detection in power systems. Synchrophasor data from controlled simulations will be used to detect events such as single line ground faults (SLG), Line to Line (L-L) faults and three phase faults (3Φ) and classify events accurately.

A typical PMU can measure about 18 parameters including phase voltage, phase current, frequency, sequence voltage, sequence currents, etc. Synchrophasor data are meant to be processed as a vector rather than as a scalar for synchronization purposes. The dimensionality of synchrophasor data exponentially increases as PMU deployment increases. The performance of machine learning algorithms can seriously be impaired because of "curse of dimensionality" associated with synchrophasor data. A number of dimensionality reduction algorithms are being used in a variety of applications, but most of them are applied on data already available. Very few dimensionality reduction algorithms are designed to work with data streams. Most algorithms are not designed to be run in real-time. Online dimension reduction will enable the processing of synchrophasor data in a less computationally extensive manner and to be more efficient with memory without the loss of much information. It is a real time compression of synchrophasor data. In this work, my objective is to use real time dimensional reduction algorithms for synchrophasor data so that the storage, transfer and analysis of data become computationally efficient without trading off the underlying information.

A variety of measurements, such as zero sequence current/voltage, positive sequence current/voltage, negative sequence current/voltage, phase (A, B and C) current/voltage, frequency, rate of change of frequency etc are available from each phasor measurement units. These parameters being measured by the PMU are the features for a classification problem. As the number of deployed PMUs increase, the number of features under consideration for a classification problem increases rapidly. All the measured parameters may not be equally important for a classification problem, depending upon the target classes (faults/contingencies).

Contributions

The following contributions result from this dissertation

1. Ascertain effectiveness of data stream mining algorithms for event detection in power systems.
2. Develop and quantify efficacy of an online feature reduction algorithm based on principal component analysis for real time compression of synchrophasor data.
3. Develop methodology for using all data points from phasor measurement units for decision making without exceeding memory and computational limitation of computational resources.
4. Identify and analyze correlation between measured synchrophasor parameters such as voltage, phase angle and current.

Dissertation Outline

This dissertation is organized into the following parts:

- Chapter I: This chapter introduces the problem and also discusses the basic properties of future smart grid perceived by the industry. It also discusses the motivation of this research and the specific problems that this dissertation aims to contribute to solve. This chapter outlines the objectives of the dissertation and points out the contributions of the research conducted. It concludes by listing summary of each chapter of this dissertation.
- Chapter II: This chapter is the literature review of the dissertation. Basically it has three parts: use of synchrophasor data in power systems, use of machine learning algorithms in power system and different kinds of metrics used in evaluating the performance of artificial intelligence algorithms. The problem domain of electrical power system requires the application of machine learning algorithms on highly unbalanced classes. This chapter discusses the problem and solution in evaluating algorithms in unbalanced classification problems.
- Chapter III: This chapter introduces the algorithm used to reduce dimension of synchrophasor data. It details the algorithm and discusses challenges of applying the algorithm in power system domain and proposes solution of working around the problem. Finally it presents the results and concludes with future works. This work has been published in

the proceedings of 2012 IEEE PES Transmission and Distribution Conference held in Orlando, FL from May 7 – 10 2012.

- Chapter VI: This chapter introduces data stream mining algorithm for situational awareness in power systems. It introduces hoeffding tree (aka Very fast decision tree) and adaptive hoeffding tree to process synchrophasor data for fast event detection. This chapter focuses on application of decision tree for quickly detecting events such as line to ground faults, by processing massive amount of synchrophasor data within reasonable time and statistically competitive accuracy. This chapter introduces details of the algorithms, experimental approach and evaluates the algorithm based on several meticulously selected experiments to prove the usefulness of the algorithm in power system domain. This work has been submitted for publication in IEEE Transactions on Smart Grid.
- Chapter V: This chapter introduces an offline dimension reduction algorithm that utilizes optimization of mutual information between state of power system and synchrophasor measurements. It proposes an optimization criterion that is useful for extraction of information from synchrophasor measurements. It finally presents results based on clustering of data, an unsupervised method to avoid biases of the information extraction method. This work has been submitted for publication in the proceedings of the 2012 IEEE International Conference on Data Mining.

- Chapter VI: This chapter is the final conclusion and discussion of future works that can be done to extend the work of this dissertation. It summarizes the results obtained from the methodologies used and discusses the applications of the algorithms to process the synchrophasor data for situational awareness in reasonable amount of time without compromising the accuracy of machine learning algorithms to be used in next generation smart grid.

CHAPTER II

LITERATURE REVIEW

Synchrophasor technology can portray the dynamic behavior of a power system in real time providing situational awareness information for operators. It can have an impact both on online operations and offline network analysis [9]. The spectacular growth in technologies in terms of computational power and communication technologies has supported the realistic feasibility of real time, wide area monitoring of a power system [28].

As synchrophasor technology is maturing and getting ready for wide scale deployment in North American power grid, a number of researchers have been working towards application of synchrophasor data for improving real time operations; post event forensic investigation, stability monitoring applications etc. Synchrophasor technology provides valuable information about the stability of the grid in time aligned vectors. This information is not available with traditional SCADA systems depriving operators wide area visibility [9].

Phasor data can have several applications which can be broadly classified into following categories [9]

- Decision Support System for real time grid operations
- Applications for system planning, event analysis and model validation

- Control applications for automated control actions based on wide area information in addition to local measurements

Phasor Data Use for Real-Time Operations

Synchrophasor measurement units provide a wide visibility of the power system which was not available with SCADA system, thus, opening new doors for numerous opportunities to use phasor data to improve operations of the power system. “It improve operators’ ability to see and understand what is happening on the bulk power system, anticipate or identify potential problems, and identify, evaluate, implement and assess remedial measures” [9] . The conversion of phasor data into actionable intelligence without information overload is a huge area of research in the power system research community [9, 29].

Phase angle differences between two points provide a measure of power flow. Instability cannot be measured by simple threshold mechanisms as voltage magnitude because it varies widely depending upon system topology. Reference [28] proposes an algorithm to detect fast separation of phase angles among critical areas. The proposed algorithm works only using synchrophasor data without knowledge of relay status. Reference [30] discusses a scheme that can use synchrophasor angle difference as a key signal to increase allowable power stability margins.

At Washington State University, researchers have developed a real time oscillation monitoring system (OMS) for detecting the emergence of small-signal instability related events in large electric power systems [31, 32]. OMS is designed to work on a specific set of rules without any human intervention during analysis. The monitor will thus issue operator alerts or control triggers whenever the damping levels of the electromechanical modes of oscillation go below preset thresholds [31].

In [33] a synchrophasor based online voltage stability index (VSI) has been proposed that predicts steady-state voltage stability limit. The proposed method simplifies a large system behind a load bus into a single source and a single transmission line using time-synchronized phasor measurements and network parameters. It provides voltage stability margin of each individual load bus in an informative format and identifies the load bus that is most vulnerable to voltage collapse [33]. This method requires a lot of "offline" manipulation of system topology before it can be used in real time. In [32], a real time estimation of static system stress margin is provided purely by monitoring the PMU measurements. The system stress is classified as normal, alert and alarm condition based on calculated stability margin. This method creates a library of critical states and their stress levels as training data for a support vector machine (SVM) in limited numbers using the generalization capability of SVM to identify stress levels of unseen critical states [32].

In [23], synchrophasor data is used for voltage security assessment using decision trees. Generator VARs and angular differences are considered in a decision tree model as both are considered as good indicators of voltage security status. A stressed power system is characterized by widening angular separation of bus voltage angles as it moves towards voltage insecurity [23]. In [34], synchrophasor data are used to detect long term voltage stability. Simulation of plausible raw PMU or state estimator outputs data are used by adding noise to bus voltages provided by snapshots of detailed time simulation. Each snapshot is fitted into an extended set of equilibrium equations from which efficient sensitivity analysis is performed [34].

Real Time Dynamics Monitoring Systems (RTDMS) have been developed as a framework for phasor based applications that can be used by operators, reliability coordinators and engineers to simulate dynamic performance of power system in real time. It provides a suite of applications ranging from post event analysis to real time monitoring of system metrics related to grid stress, dynamics, and the power systems proximity to instability such as to enable utilization of time-synchronized phasor measurements for reliability management with overall objective of accelerating the adoption and fostering greater use of the technology within North America [35].

Several applications of synchrophasor data for monitoring real time operations of power system has been proposed and is a very active area of research around the world. In this section, a few techniques have been discussed to illustrate the state-of-the-art. Many applications of synchrophasor include machine learning algorithms as a part of decision making process. Some applications use a machine learning algorithm as a sole decision maker while several of them use it in conjunction with analytical methods. In this research work, the main focus is to use machine learning algorithms. The next section will discuss machine learning algorithms being used in power systems. It is not limited to synchrophasor data rather a wide variety of applications are included.

Machine Learning in Power system

Machine learning is a branch of artificial intelligence which seeks to imitate human learning processes using inductive inference methods through observed examples. The main objective of machine learning algorithms is to extend knowledge learned from training samples to make prediction on unforeseen inputs [36] (i.e., generalization). Machine learning algorithms learn a function, often non-linear and complicated, mapping inputs to output as shown in Figure 4. The training samples of any machine learning algorithm are assumed to be an inclusive set of general properties, so that a general representative model of a system could be built.

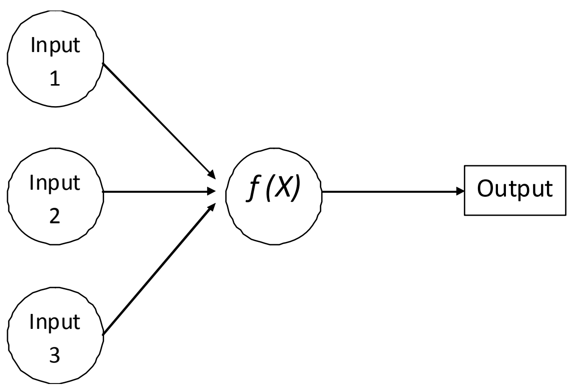

Figure 4 Mapping of input to output by machine learning

Machine learning application in power system have been explored by the research community since the 1960s [10]. Artificial intelligence (AI) techniques in conjunction with traditional analytical methods comprise a significant number of publications in power system community. Artificial Neural Networks (ANN), Support Vector Machines (SVM), Expert System (ES), Fuzzy System (FS), Genetic Algorithms (GA) have been used on a wide variety of applications in power system such as planning, security assessment, power generation optimization, Unit Commitment and Economic Dispatch [21, 37-40].

The power system has always relied on mathematical calculations/simulations and historical data for accurate prediction and forecasting. The industry still heavily relies on preventive model of operation. Generally, the power system operates at a point where it can sustain predicted contingencies without any undesired outcome. The safe point of operation is generally more expensive to operate on than a relatively “unsafe” point of operation [41]. With the integration of renewable energy sources such as wind, solar etc, the grid is destined to be pushed towards instability as per current standards [42].

The deployment of Phasor Measurement Units (PMU) has made real time wide area surveillance of power system possible. Instead of existing preventive model of operation, the energy industry is exploring possibilities of corrective model of operation in which corrective actions are initiated when system is detected heading towards instability. In this model, power system can operate closer to its capacity; utilizing existing infrastructure such as transmission lines, generators etc, instead of building new ones, thus, saving billions of dollars.

Real time monitoring of power system requires fast event detection technology, which can process synchrophasor data in real time to predict events. The traditional analytical methods do not provide enough speed in detection of events; though they provide excellent accuracy. Machine learning techniques are designed to emulate complex mathematical systems, such as power system, within reasonable latency times between input and output. Many researchers have used machine learning algorithms to replace simulations for fast decision making.

Power Quality Assessment

Power quality is a major concern for electrical utilities. It affects efficiency and life of consumer electronics. With more and more electronics device becoming part of customers day to day life, many utilities have started to have a dedicated power engineer to tackle issues of quality of power delivered to customers [43]. Electric Power Research Institute (EPRI) estimates that the US economy is annually losing between $15-$24 billion dollars to power quality issues [44]. Researchers have used several Artificial intelligence techniques to identify disturbances to help utilities to improve power quality issues such as voltage sag, voltage swell, voltage interruption, frequency deviation etc [11, 43-46].

The automatic classification of Power Quality (PQ) disturbances has been accomplished with several types of machine learning techniques. In [43] Artificial Neural Network (ANN) classifier is used to identify disturbances such as high/low frequency capacitor switching, impulsive transients etc. The feature extraction of a disturbance is done with the help of wavelet transform (Figure 5) before it is processed via a classifier [11, 44, 45]. In [46], Scales 1-5 of wavelet coefficients are used as input to three teams of five ANNs each. The final decision on classification of the power quality disturbance is a combination of decisions made by each team [46]. Neural network classified events such as high frequency capacitor switching, low frequency capacitor switching, impulsive transients etc with high accuracy in frequency domain [46]. In [18], ANN in conjunction with wavelet transform classifies transient phenomena to distinguish between internal fault and magnetizing inrush current in power transformer. The spectral characteristics of the waveform are obtained with wavelet transform and a trained ANN is used to help in power transformer protection [18]. Most of the techniques described for power quality classification seem to have used wavelets as the feature extractor before processing it using machine learning technique. As transients are very difficult to be analyzed in time domain (default domain of waveforms), the disturbance waveforms are converted into frequency domain using wavelet transform for easy classification of the transient phenomena [11].

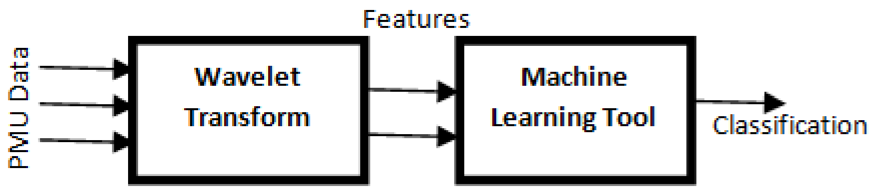

Figure 5 Pre-processing of phasor data via wavelets

In [11, 45] a different approach of using Support Vector Machine (SVM) for power quality disturbance classification is used. In [11], a SVM trained with 250 synthetic data patterns is used to classify the patterns of power quality disturbances such as voltage sag, swell, interruption, harmonics, and transients. A set of $[m(m-1)]^{/2}$ binary SVM classifiers are used to solve the multiclass classification problem. The final decision is based on the vote of each of the classifiers [11]. It is a very simple approach to solve problem of multiclass classification which is more complex to achieve using SVMs. Reference [45] uses a combination of fast fourier transform and wavelet transform for solving multiclass classification of power quality disturbance. It uses features extracted from both preprocessing methods (e.g., fundamental component, phase angle deviation, total harmonic distortion, low-frequency harmonic distortion rate are extracted using Fourier transform while energy of wavelet coefficient using two scale wavelet transform).

The learning methods such as SVM and ANN require extensive training before they are ready to classify power quality disturbances. The methods will be able to classify disturbances that they are trained on [47]. Knowledge Based Expert System (KBES) is another method of embedding the domain knowledge in terms of rules that system can infer from. KBES seeks to emulate a power engineer problem-solving process based on embedded set of rules [47].

The supervised learning approaches require an extensive knowledge base to be embedded inside the algorithm. These methods need human experts to encode knowledge known to humans in the form of training data or an experiential knowledge base. These systems are not likely to discover information that humans are not already aware of. The supervised learning methods learn non linear relationships between input and output parameters and generalize the relationship to predict to unforeseen samples. In [44] a new unsupervised learning method of Self Organized Learning ARray (SOLAR) has been proposed for classification of power quality. The multi resolution features of disturbance are extracted using wavelet transform. The features are then fed to a SOLAR for classification. SOLAR is a data driven learning method rather than knowledge driven method [44].

Fault Detection and Classification

Fault detection and classification is another of several applications of machine learning techniques applied in power system. Fault diagnosis is a way of determination of system failure depending upon measured parameters of the power system [48]. Noise, missing parameters etc. complicates mathematical system failure prediction. Machine learning algorithms can make an “acceptable” guess of output variable (fault detection/classification) even in adverse conditions such as missing data and/or noisy data. The fast decision making capability of AI techniques make them extremely desirable in real time applications.

In [49] Artificial Neural Network (ANN) has been used for classifying faults after differentiating normal maintenance operations from faults. It is a two step process of detecting and classifying faults. The detection step uses Discrete Wavelet Transform (DWT) using current wavelet coefficient energy. If no fault is detected, then no classification is performed. If a fault is detected, then voltage and current samples related to fault clearing time are analyzed for classification of the fault. The moving window of five consecutive voltage and current samples are fed as input to ANN for classification of fault [49]. Various architectures, such as radial basis neural network, are used to identify patterns in voltage and currents for classification of faults (e.g., Single Line to Ground (SLG) faults, line to line (LLG)) and has been utilized and proven to be useful [19].

Self Organizing Map (SOM) is an unsupervised alternative for fault classification. SLG, LLG, three phase bolted faults are classified correctly in [48] [20] using SOM. The requirement of large training set is eliminated using unsupervised learning method. A wide variety of operating conditions can be emulated in training learning methods rather than overtraining on the same set of operating conditions. The variation of load, incidence angle for faults, location of fault, load angles, fault resistance source impendence can be used for training [20]. If the supervised learning is used, then each of these set of data has to be hand labeled before training which may be not be practical to achieve in dynamic systems. However, unsupervised learning methods such as SOM can learn patterns in data by itself when implemented in real world [20].

Stability Assessment

Stability prediction of power system is one of the most useful features that power engineers may desire to have. There are several machine learning techniques used for predicting stability in power systems. Stability is ability of power system to remain under equilibrium under normal condition and to regain state of equilibrium after being subjected to severe disturbance [41]. Traditionally, stability means ability of maintaining synchronous operation of synchronous machines for generation of electrical power. Instability may also be encountered without losing synchronism, such as in Voltage Instability [41]. Voltage stability is ability to provide an acceptable voltage on all buses of power system after a disturbance [41]. A number of machine learning techniques have been studied for real time stability analysis of power system [14-16, 24, 50, 51].

Real time security assessment had not been feasible because of latency encountered in mathematical calculations [14, 15]. Support Vector Machines (SVM) has been studied in [14, 15] to predict transient stability of power system. Single line attributes such as machine angle, machine speed, machine terminal voltage, electrical active output power, electrical reactive power output, derivatives of machine angle and speed to time has been used as input parameters. SVM is used to differentiate between fault and post-fault measurements [14]. In [15], a comparison between neural network method and SVM is studied for transient stability analysis. SVM is more appropriate for using in a large power system where number of parameters to be monitored is exponentially large. SVM performs better when dimension of input parameters is large [15]. In addition, interpretability of the SVM results is better than that of weights of neural networks [15].

Selection of machine learning method depends upon properties of a method that best suits a problem. Decision trees are one of the most popular methods for stability assessment in power system. In [24], a single decision tree is used to predict all fault locations in network of New England 39 bus test system. Bus faults and line faults ranging between 1-8 cycles are studied. Sensitivity of machine learning techniques towards variation of operating conditions is one of the issues while using machine learning algorithms in power system. In [24], a range of fault durations from 1 to 10 cycles are simulated for transmission line faults in different operating conditions is used to increase robustness of decision trees. The operating points in [25] are varied by changing total load of system. In [22, 26], decision trees are used to identify critical attributes (CAs) from a set of system parameters from Phasor Measurement Units (PMUs) which are important for dynamic security assessment of power system. As shown in Figure 6, critical attributes are identified as A, B and C. The more important they are for classification, the more closely they are to the root of the tree. An insecurity score can be calculated for each path from root to leaf, if the score exceeds a limit and if the path associates with a probable contingency then a corrective measure can be taken by the operators [22]. In order to incorporate changing operating conditions in power system, periodic update of decision tree (DT) has been proposed in [22].

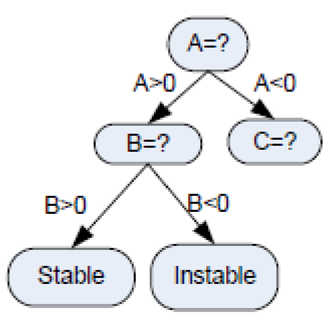

Figure 6 Example of a decision tree

A single decision tree explores more of a limited search space than that done by other methods of machine learning [52]. Decision Trees are also prone to variability in data; if the first splitting variable is chosen incorrectly, then the entire structure of the tree could be fallible to make correct predictions [52]. In order to integrate easy interpretability of decision trees (DT) and make it more robust, [52] proposes a new method of using an ensemble of decision trees (random forest) by randomizing the split at each node of a tree as a committee of experts. The random forest also allows ranking security levels of the classified instances based on sample probability estimation [52].

Decision Trees (DT) only can provide a classification of operating point as secure or insecure. Instead of providing just classification of security, Regression Trees (RT) has been proposed in [53] to provide severity of classification result by improving on algorithm proposed in [22]. In [53], RTs are used for assessment of Voltage Magnitude Violation (VMV) and Thermal limit Violation (TV) caused by N-1 contingencies. Each RTs maintains a severity score on the terminal nodes, a larger score denotes a larger elements and a more severe VMV/TV [53].

Critical clearing time (CCT) is the maximum time that a system can withstand a fault before the fault is cleared. If a fault is cleared within CCT, then system remains stable; otherwise it slides towards instability. The estimation of CCT is very important for protection and control of power systems. CCT is a complex function of operating conditions, fault structures and post fault conditions requiring complicated integration of variables for estimation [54]. In [30], a feed forward neural network has been designed to estimate CCT to replace complicated mathematical methods, which would be helpful for real time security assessment. Reference [55] takes a different approach in employing neural networks (ANN) for predicting dynamic stability by using angle of instability as predictive measure. The algorithm divides a power system into critical areas. A center of angle for the entire system is calculated based on phase angle measurements by PMUs at each generator bus. A neural network is designed to predict the angular instability of entire system based on the divergence of individual phase angles measured at each generator from center of angle of entire system (See Figure 7) [55].

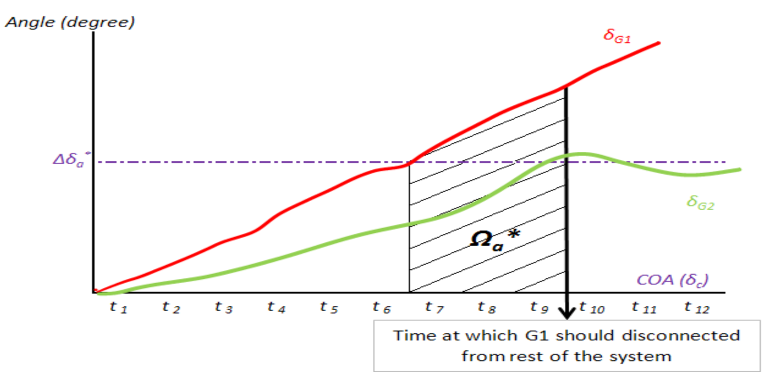

Figure 7 On-line determination of angle of instability [55]

Voltage Instability may be caused by large disturbance, increase in load demand or change in system conditions [41]. An operator needs to monitor voltage profile of system to prevent a voltage collapse. Although voltage instability is a local phenomenon, it may be a consequence of series of events accompanying voltage instability over a wide area. A possible blackout can be averted if a corrective action could be taken to prevent the cascading of the events [41, 56]. A real time monitoring of the voltage stability of power system is an important aspect of wide area monitoring of system. In [56], a real time estimation of voltage stability using PMU data has been proposed using decision trees. An operating point can be classified as Secure or Insecure in terms of voltage stability margins. Even if a system is classified as secure for an operating point, an operator may be interested in determining how secure a system is or distance to voltage instability. The stability margin may be additional real power that can be carried by the system before entering instability. In [56], a number of scenarios such as single line outages, double line transmission outages and generator losses, are tested on different loading conditions to emulate a real power system operation to construct a robust decision tree (DT) that is less prone to dynamicity in power systems. PMU measurements such as Active power flow/injections, reactive power flow/injections, voltage magnitudes and phase angles are used as input for determination of voltage stability [56].

It would be impractical to cover all possible operating points even for a trivial system. If a new operating condition arises, then decision tree/neural networks have to be retrained forgetting most of the knowledge acquired in prevision training. This process is both time consuming and inefficient for dynamic systems such as power system. In [57], a new method of incremental learning based neural network is proposed for transient stability analysis which can learn new operating conditions without forgetting previously learned knowledge. The neural network is trained offline to learn characteristics of a system, but unlike conventional neural network, the incremental neural network can learn knowledge of new training samples while predicting based on previously acquired knowledge as shown in Figure 8 [57].

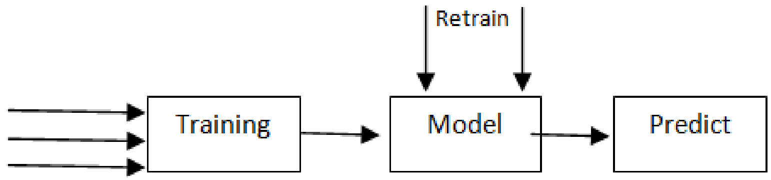

Figure 8 Incremental learning method

In [51], Kohonen map has been introduced for dynamic security assessment of power system. Voltage and power (Active/Reactive) are used for clustering operating points to estimate stability index of system. In [17], a Kohonen map is used to classify an operating condition as safe, critical and unsafe based on active and reactive power of power transmission lines.

Evaluation Metrics

Traditional data mining algorithms are widely based on the assumption that classes in a problem are uniformly balanced [58]. Unbalanced class distribution is characterized by there being many more instances of some classes than others. In traditional data mining algorithms, balance in classes is maintained during training phase of an algorithm. Appropriate number of training samples assigned to each class is identified. However, the balance between classes cannot be maintained in a data stream because of lack of knowledge about the data stream ahead of time. In power system operations, disturbance data is a very small percentage of normal operation data, even a naïve majority classifier can produce large accuracy if assumption of balanced classes is used [27]. This approach is not able to correctly measure the actual performance of data mining algorithms [58]. It is not possible to maintain balance in classes in stream classifiers. In this section evaluation measures of classification performance to handle class unbalance problem will be explored.

Some of the evaluation measures, such as recall, precision, F-measure, G-mean and Receiver Operation Characteristic (ROC) Curve Analysis, for unbalanced data problem. F-measure is used when only the detection of minority class is important. It considers both the precision and recall of the algorithm as shown in the following [58, 59].

G-Mean is used when the performance on both classes are important. G-mean as the geometric means of recall values of every classes [58, 59]. It is calculated using Equation 1.

$$\mathrm{G\ mean} = \sqrt[k]{\prod_{i=1}^{k} \mathrm{recall}^{i}} \qquad \text{Equation 1}$$

Receiver Operating Characteristic (ROC) is another popular method to evaluate learners of the unbalanced matrix. It is a plot of true positives and false positives of classified examples evaluating the pros and cons of the classifier under study [59]. The area under a ROC curve (AUC) provides a single measure of a classifier's performance for evaluating which model is better on average [59].

In [60], Kappa Statistics has been proposed as a new measure for evaluating the performance of data mining on unbalanced classes. Kappa statistics normalize the classifier's accuracy with the chance accuracy because there may be a possibility of being accurate solely by chance when predicting on unbalanced data. It can also comprehend potential drift in class distribution [60]. The calculation is based on the difference between how much agreement is actually present ("observed" agreement) compared to how much agreement would be expected to be present by chance alone ("expected" agreement) [61]. According to [60], kappa statistics can be defined as

"Consider a classifier h, a data set containing *m* examples and *L* classes, and a contingency table where cell C_{ij} contains the number of examples for which h(x) = i and the class is j. If h(x) correctly predicts all the data, then all non-zero counts will appear along the diagonal. If h misclassifies some examples, then some off-diagonal elements will be non-zero,

$$\rho_0 = \frac{\sum_{i=1}^{L} C_{ii}}{m} \qquad \text{Equation 2}$$

$$\rho_c = \sum_{i=1}^{L}\left(\sum_{j=1}^{L}\frac{C_{ij}}{m}\sum_{j=1}^{L}\frac{C_{ji}}{m}\right) \qquad \text{Equation 3}$$

In problems where one class is much more common than the others, any classifier can easily yield a correct prediction by chance, and it will hence obtain a high value for p0 (See Equation 2). To correct for this, the κ statistic is defined as follows" [60]. Table 1 illustrates a basis for interpreting kappa statistics.

Table 1 Interpretation of Kappa Statistics [61]

	Poor	Slight	Fair	Moderate	Substantial	Almost Perfect
kappa	0	0.2	0.4	0.6	0.8	1

$$\kappa = \frac{\rho_c - \rho_a}{1 - \rho_c} \qquad \text{Equation 4}$$

Summary

This chapter covered a different application of synchrophasor data for real monitoring the real-time operations of power system. A variety of applications such as power oscillation monitoring, frequency stability, voltage security assessment, have been studied utilizing just synchrophasor data for a wide area monitoring of power system. Similarly, a few applications have been designed to improve state estimation using synchrophasor data.

The deployment of phasor measurement units (PMU) can help describe the dynamic behavior of power system because of high data acquisition rate, but the major advantage of synchrophasor technology is time synchronized data which can give a snapshot of power system at any given time. Synchrophasor technology also brings the challenge to visualize the dynamic character of power system inside control room so that correct information is delivered without information overload. It also brings on the challenge of information extraction from continuous stream of synchrophasor data.

Several machine learning and pattern recognition algorithms such as support vector machine, artificial neural network, knowledge based expert system, etc have been studied for information extraction in power system since 1960s. There are several applications of Decision Trees (DT) that has been specifically designed for security assessment using synchrophasor data.

The main problem with the machine learning algorithms is that they have been designed to operate in a scarcity of data. They make multiple scans of the same training data, build a model and finally start making predictions. If the same algorithms are used for continuous data streams, such as synchrophasor data stream, the model created would be too large to accommodate in memory and the latency in predicting using the model would not be suitable for real time applications [27]. Data mining algorithms have to be adapted towards handling possibly never ending data without exceeding memory and latency requirements for real time operations. In this research work, stream mining data algorithms will be studied for real time application, which would be a contribution in research community.

I found no publication/articles that explore correlation between parameters of synchrophasor data. Most of the literature explores the possibilities of using a few parameters such as angular divergence, active/reactive power etc individually to monitor power system operations. The work proposed in this document explores correlation between individual synchrophasor parameters, which will be a contribution towards a new dimension in synchrophasor data analysis that has not been investigated yet.

CHAPTER III

ONLINE DIMENSION REDUCTION OF SYNCHROPHASOR DATA

The electric power industry is going through the greatest paradigm shift since the discovery of electricity itself in the late 1800s [2]. The massive modernization of the industry is fueled by state of the art information technologies, an exponential increase in computational power, and power system monitoring advancements, such as synchrophasor technologies [2]. A principal component of the smart grid initiative is the utilization of massive data sets to make future grids more efficient, reliable and environmentally friendly with minimal financial burden to the utilities and their stakeholders.

Electrical systems are much interconnected systems via tie lines and control areas. Thus, a disturbance in one utility can propagate to other interconnected systems. A Wide Area Monitoring (WAM) of a power system is necessary to help ensure that a disturbance in a utility does not disrupt the operation of another [42]. A WAM system is one of the key requirements for future smart grids. GPS synchronized synchrophasor data at a high speed has made the vision of WAM attainable. PMUs can sense parameters such as voltage, current, frequency etc. of a power system typically at 30 samples per second compared to one sample per 2-4 seconds in SCADA system.

The Tennessee Valley Authority (TVA) presently handles 120 online PMUs with 3.6 billion measurements archived per day with a storage size of 36GB [62]. The amount of data is set to increase exponentially as more PMUs are brought online. The explosion of time synchronized data has brought a tremendous opportunity for researchers to view the electric grid in a never before seen before perspective. It has also brought a challenge to transmit, store, analyze and retrieve massive data efficiently.

In addition to a continuous data stream from a PMU, synchrophasor data also tends to have a large dimensionality. Table 2 shows measurement types provided by a typical industry grade PMU at any given time stamp.

Table 2 Parameters measured by two PMUs

N60	Phase A Voltage
	Positive Sequence Current
	Negative Sequence Current
	Zero Sequence Current
	Ground Current
	Phase B Voltage
	Phase C Voltage
	Phase A Current
	Phase B Current
	Phase C Current
	Positive Sequence Voltage
	Negative Sequence Voltage
	Zero Sequence Voltage
	Rate of Change of Frequency (dF/dt)
	Frequency
SEL421	Phase A Voltage
	Positive Sequence Current
	Phase B Voltage
	Phase C Voltage
	Phase A Current
	Phase B Current
	Phase C Current
	Positive Sequence Voltage
	Rate of Change of Frequency (dF/dt)
	Frequency

The electric grid requires timely information to effect control actions to minimize outages. A corrective action not taken within some critical time period cannot mitigate an escalating situation. Operators need to a situational awareness of the grid in real time, so that coordinated corrective actions can be taken. The high dimensionality of PMU data can disrupt expedient extraction of information from the high speed synchrophasor data stream.

In this chapter, we will discuss a method of dimensionality reduction of synchrophasor data utilizing principal component analysis (PCA). This method will extract correlations between measurements summarizing trends in PMU data. Transmission, storage and computation of data become less expensive after dimensionality reduction of the synchrophasor data. The algorithm discussed in this paper is an online algorithm; it can summarize data in a single scan and adapt to both abrupt and gradual changes automatically [63].

Dimension Reduction Techniques

Dimension reduction is a process of reducing the amount of data with minimal loss of the information content of the data. With advancements in data collection techniques, most areas of science and engineering are overwhelmed with the amount of data waiting to be analyzed. Dimension reduction is not a new area of study. It has been studied for a long time by researchers in statistics, computer science, machine learning, signal processing etc. There are two major areas of study in dimension reduction.

Feature Selection

The complete feature set describing a data set carries all of the information of the data. However, a subset of features can often be used to describe certain underlying trends within the data. . The feature selection technique chooses a subset of "important features" from the total set of features without altering the original data.

There are several strategies to find an optimal set of "important features". Feature selection techniques can be broadly classified into feature ranking and subset selection categories. Feature ranking methods rank available features based on parameters such as information gain or distance [64]. A subset of highly ranked features is selected as a representative set of the original data set. A decision tree is an example of the feature ranking method. Subset selection techniques evaluate a subset of data against a model and modify the model until a satisfactory subset is obtained [64]. Genetic algorithms fall in this category of subset selection.

The feature selection technique may be supervised or unsupervised. Evolutionary algorithms, such as genetic algorithms, encode domain knowledge as a fitness function [65]. Decision trees require a supervised training approach to select features. Feature selection using clustering algorithms do not require any training[66].

Feature Extraction

Feature extraction is a method of reducing dimension that extracts relevant and unique information from the data set. In this method, the original signal is mathematically modified and a new set of data with smaller dimensionality is generated as shown in Figure 9.

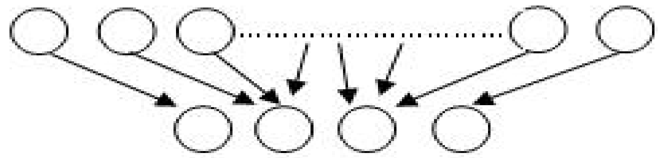

Figure 9 Illustration of dimension reduction process

The importance of a feature may depend on the application. Most Artificial Intelligence (AI) techniques, such as Back Propagation Neural Networks [67], feed forward neural networks [68], Kohonen maps [69], etc are examples of feature extraction. There are several other mathematical feature extraction processes such as Principal Component Analysis (PCA), wavelet methods, Singular Value Decomposition (SVD), etc [70].

Feature extraction processes may be either supervised or unsupervised. Supervised methods, such as neural networks, learn from training data. The characteristics of sample data are used by feature extraction methods to generalize the importance of features, while unsupervised mathematical models, such as Principal Component Analysis, wavelets, etc. use mathematical tools to extract energy representing the importance of the features.

Online Dimension Reduction

In applications where a data stream arrives at a high rate, the processing of a data set has to be done before the next set of data arrives. This property of data streams limits traditional dimension reduction techniques that require multiple scans of data. In time critical applications, such as synchrophasor data processing for situational awareness, a new approach is required to meet the latency requirement. This section will discuss online dimensional reduction techniques that can satisfy this requirement.

Reference [71] introduces a supervised dimension reduction technique known as Linear Discriminate Analysis (LDA). It uses a sliding window method in the data stream to update within class and between class scatter matrices [71]. The dimension reduction is a function of learning the mapping from higher dimensional space to a lower dimensional space. A radial basis function (RBF) has been used in [72] for the mapping. The supervised method proposes two modules. The first module generalizes the sample data using geodesic distance in data space. The dimension reduction module then uses this information to approximate the radial basis function from higher to lower dimensional feature space [72].

Principal Component Analysis

The electric grid is a dynamic system for which an infinite number of operating conditions may exist. If a supervised algorithm is used, it needs to be trained for many operating points and be able to generalize solutions for the total suite of operating points. However, tracking system operating points is itself a research problem while generating a set of training samples for these operating conditions is almost impossible to achieve. The dimensional reduction algorithm for the streaming data from the grid would be practical to use if it is unsupervised and can incrementally adapted to both abrupt and gradual changes. The memory and computation requirements also need to be minimized while processing the high speed synchrophasor data stream.

Principal components can be thought of as representing the energy of the original data and are used to help explain the variance in the data. The number of principal components representing the original data depends upon the amount of energy that has to be conserved and the first principal component explains the majority of the variance in the data with following components explaining lesser amounts of the variance

To illustrate an online dimension reduction via a PCA approach assume at time t, synchrophasor data arrives as a n-dimensional vector Xt = [xt,1, xt,2, xt,3, …… xt,n]. The synchrophasor vector is comprised of electrical parameters (features) such as frequency, voltage, current etc. There may be some correlation between these parameters in a steady state operating condition. However, the correlation gets changed during disturbances and evolves to a new correlation when the system evolves to a new operating condition. In the proposed method of dimensional reduction, correlation of electrical features will be tracked using principal component analysis. It does not require buffering of past measurements, which can be discarded as soon as new set of synchrophasor data arrives [63].

Let wi = [wi,1….wi,n] be the participation weight vector for the ith principal direction. The hidden variables yt= [yt,1…yt,k] and the projections of xt onto each wi, over time is defined as Equation 5

$$y_{t,i} = w_{i,1}x_{t,1} + w_{i,2}x_{t,2} + \ldots + w_{i,n}\, x_{t,n} \qquad \text{Equation 5}$$

Let $x'_t = [w_{1,j}\, y_t, \ldots w_{n,j}\, y_{t,n}]$ be the reconstruction of x_t using weights and principal components defined as Equation 6

$$x'_{t,j} = w_{1,j}y_{t,1} + w_{2,j}y_{t,2} + \ldots + w_{k,j}y_{t,k} \qquad \text{Equation 6}$$

This algorithm monitors and adapts the number of hidden variables (k) to achieve a desired reconstruction error $||x_t - x'_t||^2$. It also adapts the participation weight (w) to correctly summarize the original data.

Energy thresholding is the deciding factor for the number of hidden values (k). The algorithm for determining k is given as follows [63].

- Estimate the full energy E_{t+1}, incrementally, from the sum of squares of $x_{t,i}$.
- Estimate the energy $E^{'}(k)$ of the k hidden variables.
- Introduce a new hidden variable if the energy represented is smaller than the threshold and drop a hidden variable if the energy represented is greater than the threshold.

The participation weights w_i $1 \leq i \leq k$ are also updated incrementally to minimize reconstruction error. If we consider $x_{t+1} = [x_{(t+1),1} \ldots.. x_{(t+1),n}]$ as n-dimensional synchrophasor data at time t+1, the following algorithm incrementally updates w [63].

- Compute the hidden variables $y'_{t+1,i}$,$1 \leq i \leq k$, based on the current weights w_i, $1 \leq i < k$ by projecting x_{t+1} onto these.
- Estimate the reconstruction error and the energy based on $y'_{t+1,i}$.
- Update the estimates of w_i, $1 \leq i \leq k$ and output the actual hidden variables $y_{t+1,i}$ for time t+1.

w_i already maintains information about the data stream up to time t, changing w_i entirely based just on data at time t+1 can make the algorithm prone to noise. If the update of the estimate of w_i is inversely proportional to the current energy $E_{t,i}$ (defined by Equation 7) of the i^{th} hidden variable [63], then algorithm becomes less prone to noises.

$$E_{t,i} = \left(\frac{1}{t}\right) \sum_{\tau=1}^{t} y_{t,i}^{2}$$ Equation 7

Data Scaling Problem

Synchrophasor data are of different scales. Frequencies, currents and voltages vary greatly in magnitudes. The standard deviations of synchrophasor measurements obtained in our experiment illustrate this problem (see Table 3).

Table 3 Standard deviation of measurements

Measurement	Std. Deviation
Phase A Voltage Magnitude (SEL421)	1297KV
Phase A Voltage Magnitude (N60)	3758.6KV
Frequency (N60)	4.6mHz
Frequency (SEL421)	3.5mHz
Phase A Current Magnitude (SEL421)	41.67A
Phase A Current Magnitude (N60)	166.78A

PCA is sensitive to scaling of variables. The information content in features is measured in terms of variation in measurements. PCA tends to retain more information about variables having higher standard deviation than that with smaller standard deviation. Normalization of this data using the per-unit system, typically used in power system analysis, did not work. Normalization of data by calculating data parameters such as, standard score etc, overcomes this problem. The characteristics of data such as mean, standard deviations, etc. are not known beforehand in case of data stream processing thus, making the normalization of data not feasible.

The data of similar type are grouped together and processed separately for dimensionality reduction. Voltages, currents and frequencies are processed as three groups so that maximum information is retained.

Results

In this experiment, we simulated single line to ground (SLG) faults on phase A for a 4-Bus, 3-Generator system to measure the performance of the dimensional reduction technique. In order to make the experiment symmetrical to both PMUs used, the inputs to the dimension reduction have been reduced to 18 measurements as shown in Table 4.

Table 4 Inputs to the dimension reduction method

N60	Positive Sequence Current Magnitude
	Positive Sequence Voltage Magnitude
	Phase A Voltage Magnitude
	Phase B Voltage Magnitude
	Phase C Voltage Magnitude
	Phase A Current Magnitude
	Phase B Current Magnitude
	Phase C Current Magnitude
	Frequency
SEL421	Positive Sequence Current Magnitude
	Positive Sequence Voltage Magnitude
	Phase A Voltage Magnitude
	Phase B Voltage Magnitude
	Phase C Voltage Magnitude
	Phase A Current Magnitude
	Phase B Current Magnitude
	Phase C Current Magnitude
	Frequency

The plot of phase A voltage and positive sequence voltage from both PMUs is shown in Figure 10. The plot shows visually significant correlation at all times.

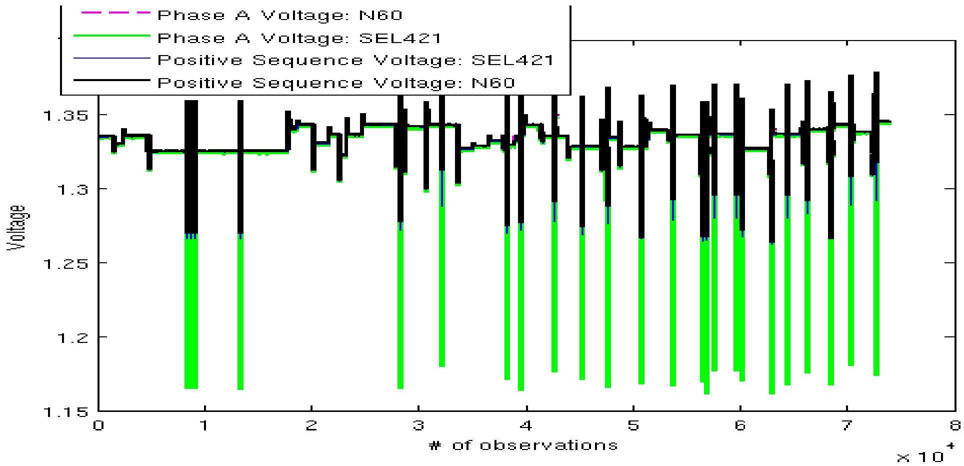

Figure 10 Phase A voltage and positive sequence voltage plot

Figure 11 shows phase A currents and positive sequence currents from both the PMUs. The currents measured by N60 are significantly higher than that measured by SEL421. Figure 12 is the plot of frequencies from both PMUs.

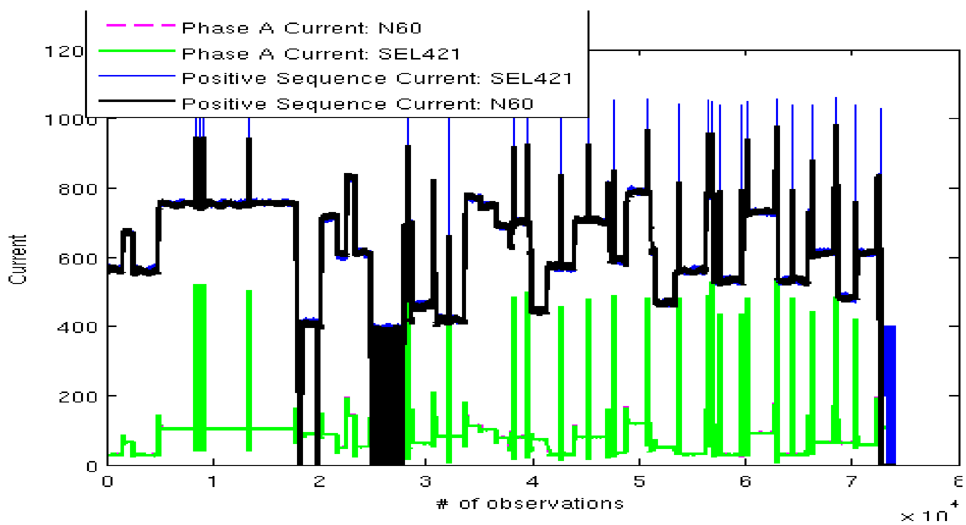

Figure 11 Phase A current and positive sequence current plot

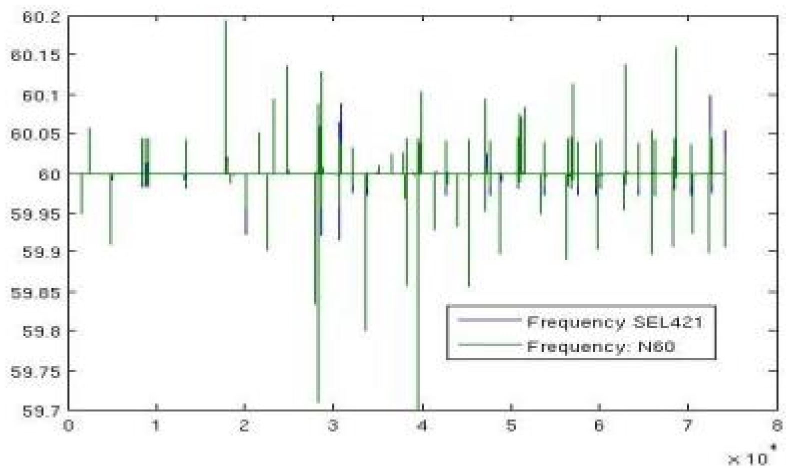

Figure 12 Frequencies plot

In order to address the data scaling issue in synchrophasor measurements, parameters with a similar scale are grouped together and dimension reduction is applied. Three groups were formed for currents, voltages and frequencies as shown in Table 5. Each group consists of measurements from both PMUs. So group 1 and group 2 contain 8 measurements each, while group 3 contains 2 measurements.

Table 5 Division of group for dimension reduction

Group 1	Group 2	Group 3
Positive Sequence Voltage Magnitude (2)	Positive Sequence Current Magnitude (2)	Frequency(2)
Phase A Voltage Magnitude (2)	Phase A Current Magnitude(2)	
Phase B Voltage Magnitude (2)	Phase B Current Magnitude(2)	
Phase C Voltage Magnitude (2)	Phase C Current Magnitude(2)	

In this experiment, 8 voltage measurements were represented by 1 principal component as shown in Figure 13 when 95%-98% of energy was retained. The performance of the dimensionality reduction algorithm was measured by analyzing the reconstructed signal as shown in Figure 14. The reconstructed signal was similar to that of original signal illustrated in Figure 10. In addition to the visual comparison Table 6 tabulates correlation coefficient and root mean square error (RMSE) of reconstructed synchrophasor data and original data for result comparison.

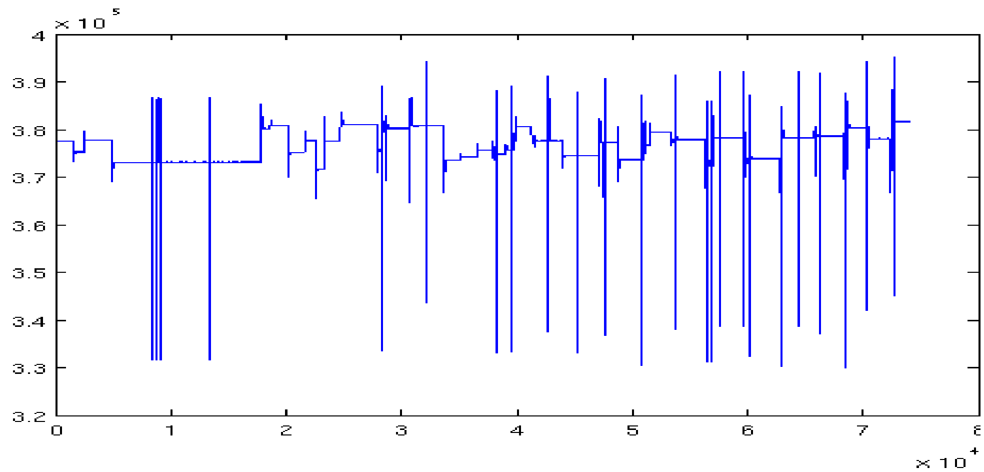

Figure 13 Principal components of voltages

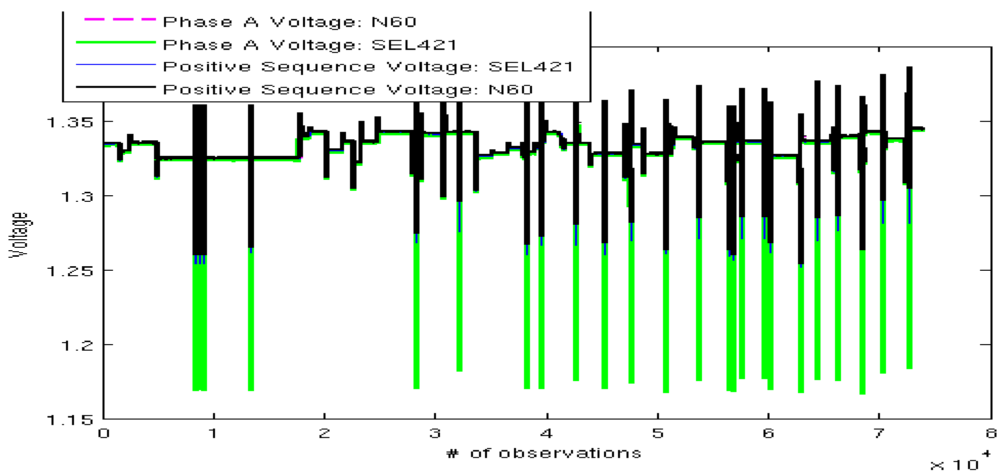

Figure 14 Reconstructed phase A and positive sequence voltages plot

Figure 15 illustrates two principal components extracted out of 8 current signals of group 2. Most of the time the first PCA was enough to capture the information, except for a few instances where the fluctuation of current was higher. The reconstructed signals of currents are shown in Figure 16 . (see Table 6 for more comparison results)

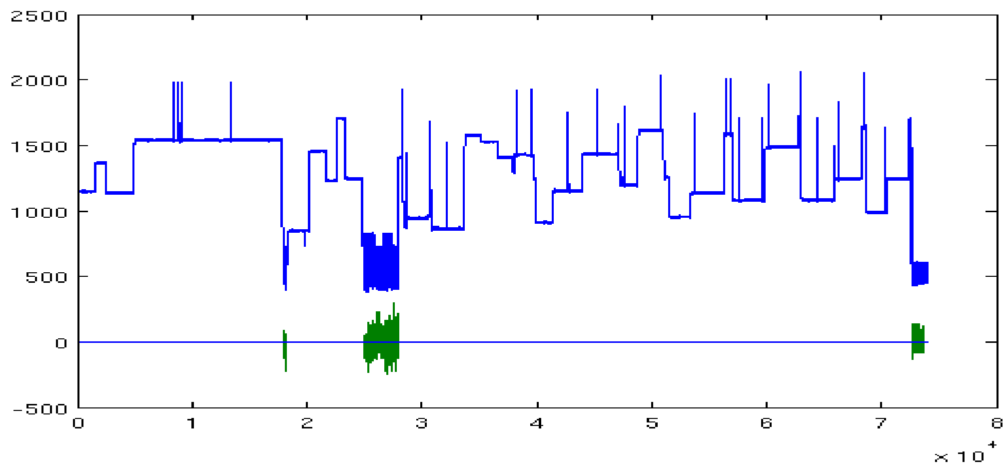

Figure 15 Principal components of currents

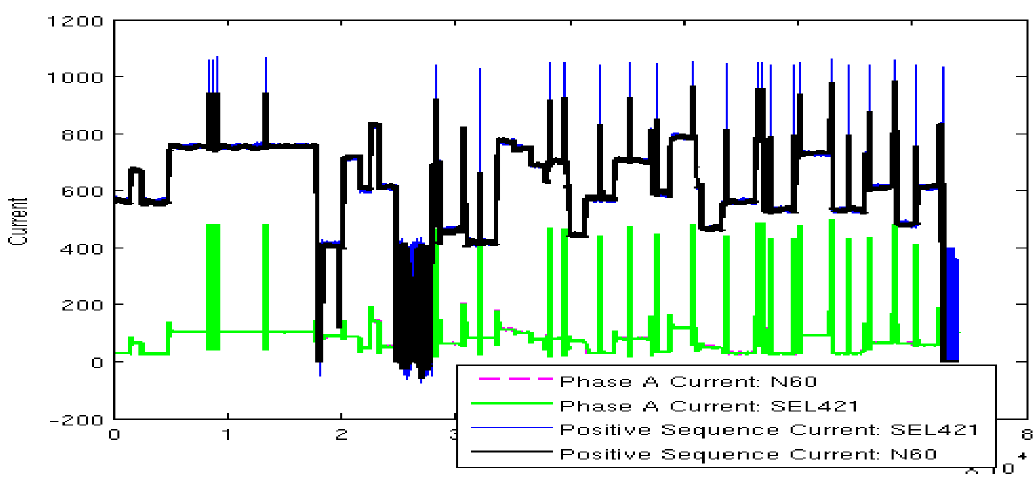

Figure 16 Reconstructed phase A and positive sequence currents plot

Figure 17 illustrates principal components of frequencies from in group 3. One principal component turned out to be enough for representing the frequencies measured by both PMUs. Figure 18 shows the reconstructed frequencies from just the first principal component. (See Table 6 for more comparison results)

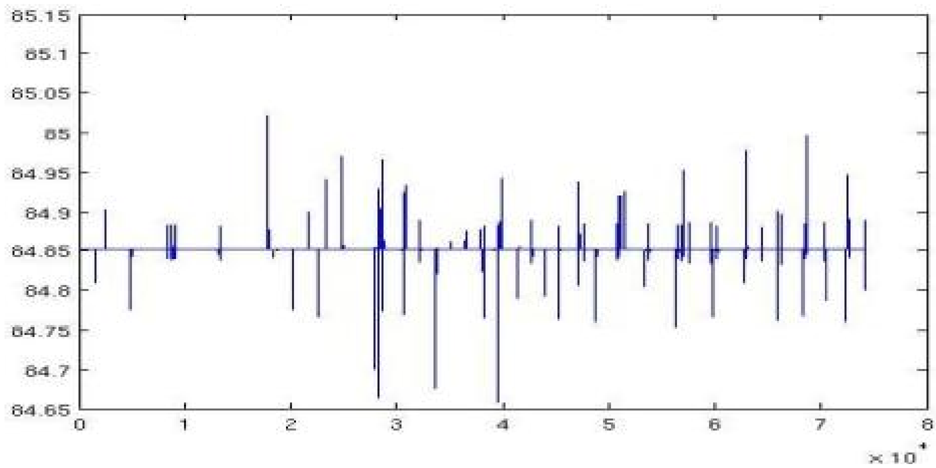

Figure 17 Principal components of frequencies

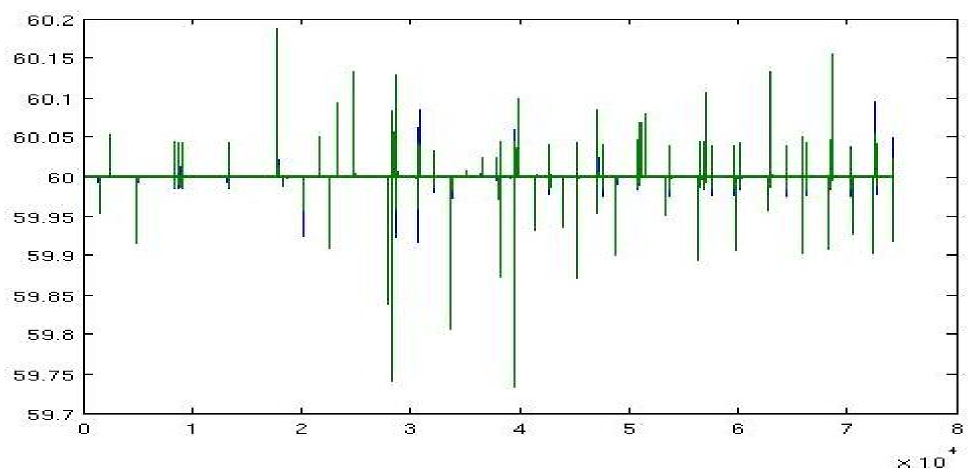

Figure 18 Reconstructed frequencies

Table 6 Correlation coefficients and root mean square error of reconstructed signal and original signal

PMU	Parameter	Corr. Coeff.	RMSE
N60	Positive Sequence Voltage Magnitude	0.9998	169.44V
	Positive Sequence Current Magnitude	0.9991	6.84A
	Phase A Voltage Magnitude	0.9986	34.56V
	Phase B Voltage Magnitude	0.9998	24.77V
	Phase C Voltage Magnitude	0.9999	41.82V
	Phase A Current Magnitude	0.9978	11.20A
	Phase B Current Magnitude	0.9994	6.54A
	Phase C Current Magnitude	0.9995	4.33A
	Frequency	0.9956	0.44mhz
SEL421	Positive Sequence Voltage Magnitude	0.9993	34.37V
	Positive Sequence Current Magnitude	0.9991	1.42A
	Phase A Voltage Magnitude	0.9999	19.66V
	Phase B Voltage Magnitude	0.9979	52.62V
	Phase C Voltage Magnitude	0.9972	59.46V
	Phase A Current Magnitude	0.9985	2.27A
	Phase B Current Magnitude	0.9993	1.21A
	Phase C Current Magnitude	0.9992	1.29A
	Frequency	0.9917	0.44mhz

The similarity index of the reconstructed signal and the original signal is measured in terms of root mean squared error (RMSE) and coefficient of correlation between original signal and reconstructed signal. The synchrophasor data reconstructed from principal components showed strong correlation with the original data, while mean square error looks acceptable in terms of the magnitude of original data.

Conclusion and Future Work

In this chapter, an online dimension reduction method for synchrophasor data was discussed. We experimented with real PMU data generated by SEL421 and GE N60 PMUs deployed on a RTDS simulation of a 4-bus, 3-generator system. The results obtained from the experiments proved that the method can be used for addressing the "curse of dimensionality" problem that machine learning algorithms may suffer when used with synchrophasor data. Though it is not a lossless method, it may be used as a preprocessing method in data analysis and data storage where absolute accuracy is not required.

Principal component analysis is a linear method of dimension reduction. The method discussed here shows a unique approach to use it in a stream of data. In the future, we plan to work towards other non linear feature reduction techniques for data streams and compare results and develop a criterion for selecting a dimension reduction scheme for synchrophasor data.

CHAPTER IV

DATA STREAM MINING OF SYNCRHOPHASOR DATA

Background

Electrical system is considered by some to be the greatest engineering achievement of the 20th century [73]. However, even 99.97% reliability of current electrical system is not enough to prevent $150 billion in losses from outages and interruptions [74]. With the increasing dependence of human enterprises on electrical energy, grid reliability has emerged to be one of the most important factors in power system operations. However, 100% reliability is a formidable challenge and is highly dependent upon the delicate balance of matching generation to load. The facts that electrical energy cannot be stored in large scale and is governed by the laws of physics makes power systems one of the most complicated systems. It is a technological challenge to maintain a continuous supply of high quality power from a generation facility to a consumer appliance.

With electric utilities responding to the nation's environmental concerns, more renewable sources, whose generation is often difficult to control, are being integrated into the grid. Revolutionary concepts such as two way flow of energy and information between grid and consumer has added a new perspective to reliability. Reliability of grid depends on constant monitoring of grid health parameters such as frequency, voltage and phase angles, and ability to make real-time decisions based on trends in monitored parameters.

Recently, there has been an increase in the deployment of Phasor Measurement Units (PMUs) which enable real time wide area monitoring of the power system. PMUs can synchronously measure operating parameters across the grid at typically 30 samples per second, compared to 1 sample per 2-5 seconds of a conventional SCADA system [75]. Such an explosion of time-stamped data in power systems has provided an opportunity to make electrical grids more reliable. Additionally, it has also brought a challenge to extract information from continuous high speed data streams. In this paper, we propose a new methodology to process PMU data based on event stream mining techniques for enhanced situational awareness in smart grids.

Introduction

Generally, power systems are designed to withstand a host of pre-determined contingencies with automatic protection and control algorithms. In the case of a rare combination of contingencies, such as the North American Blackout on August 14 2003, automatic protection systems can fail resulting a wide spread outage affecting millions of customers [76]. In transient events, reaction time is at most 100 milliseconds and therefore, automatic control equipment takes over the decision making with no human intervention in the loop. For long term stability, operators usually have enough time to run simulations and consult other operators, for making informed decisions [77]. However, there are times in between those two extremes in which operators have to use their own judgment to act on certain conditions and this often happens when there is insufficient information available to support their decision [77].

In the past, various Wide area monitoring parameters, such as phasor angles, were estimated after numerous iterations of power flow solutions, but they can now be directly measured with PMUs [78]. GPS synchronized phasor data can give operators a wide area situational awareness of the power system, which was not possible with conventional SCADA systems. The investigation of August 14, 2003 blackout pointed out that the blackout could have been prevented, if phasor data had been monitored. A number of clues surrounding the blackout were missed due to lack of situational awareness infrastructure.

With the deployment of PMUs, industry now has capability of monitoring grid health parameters in real time. However, if underlying information in high speed data stream cannot be extracted, then it is not possible for operators to make informed decisions. Typically, mathematical calculations such as power flow solution are used in power systems. However, time required for mathematical calculations to run makes it infeasible for real time situational awareness. Machine-learning algorithms, such as, Artificial Neural Networks (ANN) [79] and decision trees (DT) [24, 80-83], are being extensively studied for online prediction of power system stability based on phasor data in an actionable period of time. Conventional machine learning techniques such as ANN and DT are designed to work with a limited amount of sample data. They make multiple scans of data to build a model before making predictions.

Decision trees look promising in modeling of power systems based on phasor data [24, 80-83]. Decision trees can work with continuous data equally well as with discrete data and the results of decision trees can be interpreted by humans, which make them an ideal choice for power systems. However, the number of samples from a PMU increases exponentially as parameters being considered or number of deployed PMUs increases. For example, in a 24 hour period a single PMU produces (24x60x60x30) 2,592,000 samples for a single parameter. With the limited computational time and memory available in computer resources, this can limit the size of decision trees built using traditional machine learning algorithms. Therefore, it may be hard to accommodate a huge decision tree in limited computer memory without losing information.

One of the easiest methods to handle huge amount of data is to downsample to appropriate level. This approach is not appropriate for synchrophasor data because dynamic behavior of power system is not properly represented in downsampled data, which may even undermine the advantage of using high speed synchrophasor data. Figure 19 illustrates the disadvantage of downsampling data phasor data. Left half of the figure is PMU data at 30 samples per second (typical PMU data rate), while right half is downsampled version of same data at 0.2 samples per second (typical SCADA data rate). The details captured by PMU are lost when it is down sampled. The lost details of the synchrophasor data may be pivotal in making time critical decisions. Therefore, an algorithm which can use all data points from PMU is important to portray dynamic behavior of power systems and detect events.

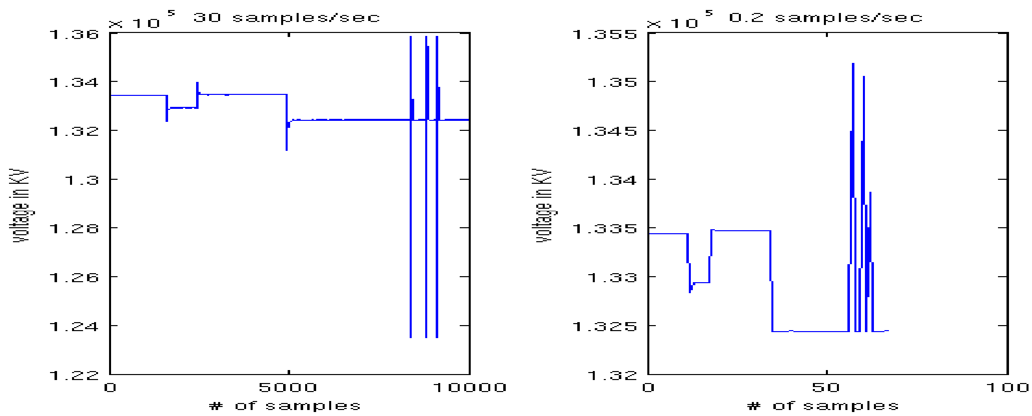

Figure 19 Illustration of dynamic behavior representation by a PMU at 30 samples per second vs. 0.2 samples per second

A new method known as data stream mining [52, 84] can extract information from high speed data streams facilitating decision-making within constraints of limited resources and time. Reference [52] builds a decision tree for data stream in limited memory using hoeffding bound to guarantee that the result obtained is as good as that of conventional decision tree. Data stream mining is a good approach for the extraction of information from PMU data stream.

Online Predictive Models for Situational Awareness

In this section, an overview of on-going research on predictive algorithms that are being used in the power system is discussed. Real time prediction algorithms, their usability and their application based on application in high speed PMU data stream is focused.

Generally, robust mathematical techniques such as power flow analysis [85] and probabilistic approaches [86] provide a reliable way to predict stability of a power system. However, in a real time grid surveillance scenario, it may not be possible to afford time lag of solutions provided by these models. As an alternative to the accurate mathematical model, researchers have been studying different machine learning techniques that will help predict events on grid in an actionable time frame.

Traditional Batch Processing of Phasor Data

Machine learning techniques learn from examples. They generalize the relationship between measured data and the state of system to predict future states of the system based on new inputs. They formulate a generalization from new data that will be applicable to most of the problem space. For example, in a handwriting detection application, a set of handwritten alphabets can be used as training data to train a system to digitize handwritten documents. Traditional batch processing machine learning techniques assume that all training data are available simultaneously. They make multiple passes on the training data and adjust themselves to create a general predictive model.

As previously stated, decision trees have been extensively used for situational awareness using phasor data. The predictor decision trees are created offline using historical phasor data, identifying critical attributes (CA) and their thresholds among several measurements from PMUs. In Figure 20, attributes A, B & C are identified as critical. The more important they are for classification, the more closely they are to the root of the tree. The path from the root of a decision tree to the leaves determines the classification of the event where leaves store the classification. Reference [82] goes one step further and uses a committee of 210 decision trees (Random Forest) to predict the dynamic system stability.

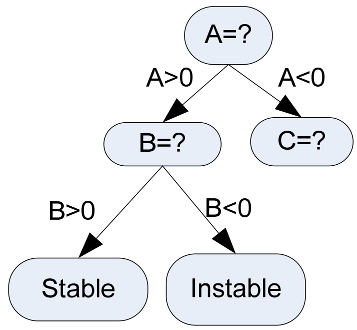

Figure 20 Example of a decision tree

When input to conventional machine learning algorithms is a data stream, it has to be stored (e.g., in a database) before an algorithm is applied. This is done to ensure simultaneous availability of training samples, as shown in Figure 21. In case of the phasor data stream, it may be possible to store data, but it may not be practical to go through massive stored data in order to predict a result in the time available for making a decision. The model created for continuous synchrophasor data continuously grow in size making it impossible to store in available memory without losing information.

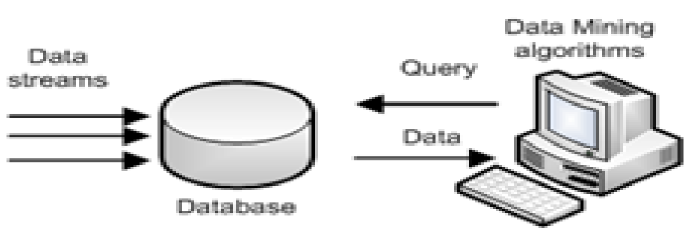

Figure 21 Batch processing algorithms for data streams

Event Stream Mining of Phasor Data

Since PMUs generate continuous streams of data, decisions have to be made before a new set of data arrives (see Figure 22). The typical data rate of PMUs is 30 samples per second which results in any parameter having more than 2 million daily samples. Parameters such as, phasor angle, reactive power, voltage magnitude etc, must be monitored from multiple PMUs to facilitate situational awareness. Therefore, a new approach is needed to handle such a massive amount of data in a limited memory and with limited computational power.

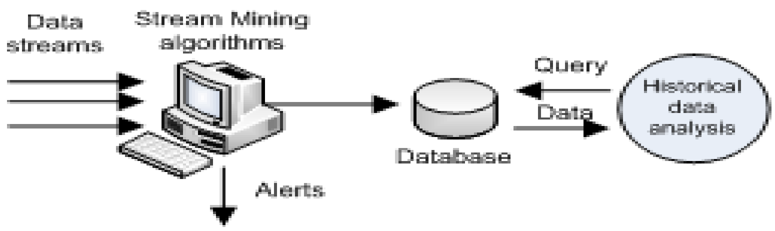

Figure 22 Event stream processing

Stationary Data Stream Mining

Stationary data stream mining algorithms assume that data streams are not evolving. The distribution of data on stream is same all the time. Data Stream mining is a relatively new field of study. It is useful in systems such as, Cyber Security [84], financial monitoring [85], homeland security [86] etc., which generate huge amounts of data in short periods of time, like PMUs.

Hoeffding Trees

Domigos and Hulten introduced Hoeffding trees in [87], which is one of the pioneer works in the area of massive data stream mining. The Hoeffding tree induction algorithm builds a decision tree by scanning the incoming data stream only once. There is no need of storing the data as in traditional decision trees. The tree itself holds sufficient statistics in its leaves to grow the tree and also to make classification decisions of incoming data.

Each node in a decision tree contains a test for an attribute, and the branch to follow after the node depends upon outcome of the test. Each leaf contains a class prediction. Classification problem works depending upon a series of such tests at each node from root to leaf. A decision tree is learnt by continuous replacement of leaves and selection of thresholds and test attribute at each node. A heuristic is needed to select attribute to be tested at each node. The most common heuristic is the information gain (G), which is a measure of discriminative power of each attribute [87]. The number of samples (λ) to be used at each node to be scanned before calculating information gain is determined using hoeffding bound (see next section).

If X_a and X_b be the two PMU measurements with two highest G calculated after seeing λ examples at a node. Let $\Delta G = G(X_a) - G(X_b) \geq 0$ be the difference between information gains, then given a desired δ the hoeffding bound guarantees that X_a is the correct choice for the split with probability 1-δ if λ samples have been seen at this node and $\Delta G > \varepsilon^2$. An algorithm for splitting a node l is as follows.

1. Create synchrophasor vector (X: C) from measurements from each time-stamped data.
2. For all training examples
    a. Update sufficient statistics in leaf node (l)
    b. Increase n, counter that tracks number of examples seen.
    c. If n == λ, then

i. Compute G for each parameter and let X_a and X_b be two attributes with highest Gs.

ii. If $G(X_a) - G(X_b) > \varepsilon$, then replace l with an internal node that splits on X_a

iii. Initialize all branches of the split with sufficient statistics.

There are several strategies that are used to prevent the size of tree from getting out of bounds as explained in [87].

Hoeffding Bound

The single most important feature of decision trees is to split a node. The effectiveness of attribute selection to split node determines the accuracy of the decision tree. Criteria such as Gini index and information gain are used for selecting attributes and in determining the "Goodness" of a resulting tree [87]. The calculation of information gain is slightly more complicated in data stream mining than in traditional data stream mining because of the unavailability of simultaneous training data to the algorithms. Domigos and Hulten proposed a criteria known as Hoeffding bound which guarantees statistically the same decision for stream mining as that with traditional batch processing algorithms [87].

The Hoeffding bound states that with probability 1-∂, the true mean of a random variable of range R will not differ from the estimated mean after n independent observations by more than:

$$\in = \sqrt{\frac{R^2 \ln\left(\frac{1}{\partial}\right)}{2n}} \qquad \text{Equation 8}$$

This bound is useful because it holds true regardless of the distribution generating the values, and depends only on the range of values, number of observations, and desired confidence. A disadvantage of this approach being so general is that it is more conservative than distribution-dependent bounds [52].

Evolving Data Stream Mining

Generally, data streams change over time, which diminishes the relevancy of built model to make future decisions. Most of real time applications are dynamic; stream mining algorithms has to constantly be adapting itself to changing distribution of data to make relevant decisions as shown in Figure 23. Thousands of customers switching their electrical appliances on and off, opening and closing of relays, and breaker operations in response to contingencies make power system a very dynamic system. Adaptability of data mining algorithms ensures that the information extracted from data is accurate for the current situation. In this section, we will discuss data mining techniques that are being researched for the evolving data stream.

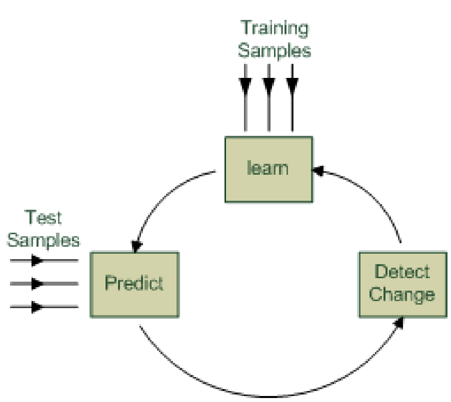

Figure 23 Adaptive algorithm for evolving stream

Hoeffding Window Trees using ADWIN

Hoeffding Window Tree is any decision tree that uses Hoeffding bounds and maintains a sliding window of instances. The algorithm maintains detectors at every node that will flag changes. It creates, manages, switches and deletes alternate trees. A change detection algorithm ADWIN has been proposed in [88] which has been used in automatic detection of change and adapts the Hoeffding tree to the current rate of change. ADWIN is a parameter free and assumption free algorithm which makes it easier for users to implement without needing a priori knowledge of characteristics of data stream.

Hoeffding Window Tree is any decision tree that uses Hoeffding bounds and maintains a sliding window of instances. The algorithm maintains detectors at every node that will flag changes. It creates, manages, switches and deletes alternate trees. A change detection algorithm ADWIN has been proposed in [88] which has been used in automatic detection of change and adapts the Hoeffding tree to the current rate of change. ADWIN is a parameter free and assumption free algorithm which makes it easier for users to implement without needing a priori knowledge of characteristics of data stream.

Hoeffding Adaptive Trees

Hoeffding Adaptive Tree (HAT) is an algorithm that adapts to changes in the data stream without requiring users to estimate the size of a sliding window to deal with the concept drift in the data stream [52]. It automatically detects the rate of change of data streams to adapt to the change of data. It places instances of estimators of frequency statistics at every node. There are several variants of HAT depending upon the estimator used.

- HAT-INC: It uses a linear incremental estimator

- HAT-EWMA: It uses an Exponential Weight Moving Average (EWMA)
- HAT-ADWIN: It uses an ADWIN estimator. As the ADWIN instances are also change detectors, they will give an alarm when a change in the attribute class statistics at that node are detected, which indicates also a possible concept change.

Experimental Approach

In this research work, stream data mining approach for detection of events in power systems is studied. Data stream mining is a new approach of artificial intelligence technique in power system application. Several experiments are performed to substantiate our proposed methods are efficient and capable of handling huge amount of data within limited resources of memory. The algorithms can predict events within reasonable time so that it can be used in real time situational awareness application in power systems. Experimental evidences will be presented to show that adaptive variant of hoeffding tree can incrementally learn changing conditions of power system, making predictions relevant to new operating point of power system. We utilize load change to simulate changing operating condition. We believe that experimental evidences support our proposal that data stream mining algorithm possess enough prospect to solve problem of mining high speed synchrophasor data to support decisions in real time.

Experimental Settings

In order to demonstrate the usefulness of event stream mining algorithms for situational awareness in power systems, we used simulations of a power system from a Real Time Digital Simulator (RTDS). RTDS is a real-time power system simulator that performs digital electromagnetic transient simulation of electric power circuits using a time step as small as 2 microseconds [89]. The real-time operation of the RTDS makes it suitable for development and testing of protection and control techniques for power systems. RSCAD is used to design the power system circuits which can be fetched to RTDS. The RTDS at Mississippi State University consists of a cubicle with two processor racks, containing eight Triple Processor Cards and two Giga Processor Cards (GPCs) [89].

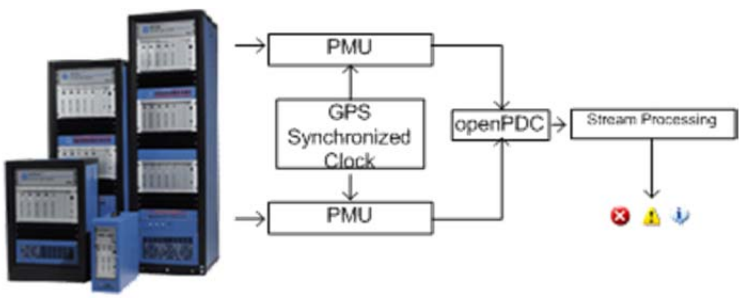

Figure 24 Data flow in experimental setup

We used a hardware-in-the-loop approach in order to make the experiment close to a real world scenario. We used two Phasor Measurement Units (PMUs): SEL421 from Schweitzer Engineering Laboratory (SEL) and N60 from General Electric. Both PMUs were synchronized with a GPS clock as shown in Figure 24 and configured at a data rate of 30 samples per second. Table 7 shows the parameters obtained from each PMU.

Table 7 Parameters measured by two PMUS

N60	Phase A Voltage
	Positive Sequence Current
	Negative Sequence Current
	Zero Sequence Current
	Ground Current
	Phase B Voltage
	Phase C Voltage
	Phase A Current
	Phase B Current
	Phase C Current
	Positive Sequence Voltage
	Negative Sequence Voltage
	Zero Sequence Voltage
	Rate of Change of Frequency (dF/dt)
	Frequency
SEL421	Phase A Voltage
	Positive Sequence Current
	Phase B Voltage
	Phase C Voltage
	Phase A Current
	Phase B Current
	Phase C Current
	Positive Sequence Voltage
	Rate of Change of Frequency (dF/dt)
	Frequency

Time synchronization enables measurements from multiple PMUs to be temporally aligned as a vector $\{x_1,x_2,x_3\ldots,x_n\}$, where ‘n’ is number of total parameters measured by all deployed PMUs. We define a synchrophasor vector (X, C), where X is a vector of *n* PMU measurements and C is discrete class indicating status of power system. The synchrophasor vector can be fetched into stream mining algorithm to identify events of power system based on signatures and trends of the measurements[90].

System Model

RTDS model of a power system is based on data provided in [91]. The power system is a four bus, three generator power system. Two phasor measurement units (PMU) are connected on Bus 1 and Bus 2 to measure electrical parameters shown in Table 7. A GE N60 is connected to Bus 1 while SEL421 is connected to Bus 2. Figure 25 shows the single line diagram (SLD) of power system model.

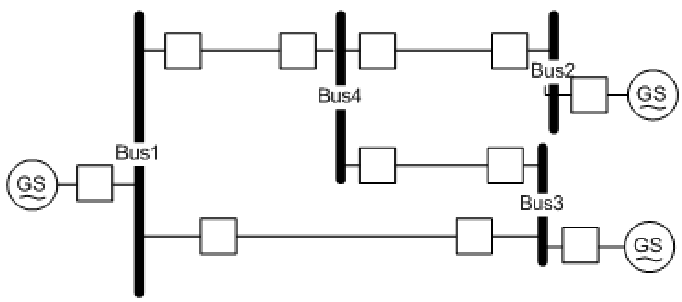

Figure 25 Single line diagram of power system under study

Result Evaluation Methods

The results of a learning process have to be evaluated on some basis to compare effectiveness of algorithms. The batch learning algorithms use the following evaluation processes.

Holdout Method

In this method of evaluation, a set of random samples are held out from training process as an independent evaluation set. An independent set is used to test the effectiveness of the algorithm on unseen samples. It is generally used when there are abundant samples in training examples [52].

Cross-Fold Method

In this method of evaluation, training set is divided into K folds. The training is repeated K times using each set as an evaluative “independent” set. The final result is average performance of the algorithm for each train/test set. It is useful when training samples are limited [52].

In order to evaluate algorithms for event data stream mining, the following method has been used in this dissertation.

Interleaved Test-Then-Train

In this method of evaluation, a sample is used for testing before it is used for training the model. The accuracy is incrementally updated. Also, the algorithm is tested on samples it has never seen before. It makes very effective use of training samples for testing. The downside of this approach is that there is no distinction between training and testing time [52].

Evaluation Measures

Several parameters can be defined to measure the performance of algorithm. Basically there are three areas of performance that we are interested in processing synchrophasor data: how accurate is the classification, how fast algorithm runs (latency) and how efficiently memory resource is utilized by algorithm. The following points give an insight on details for performance measure that have been considered for synchrophasor data processing.

Accuracy Measure

In power systems, normal data are more common than events. Events (such as single line to ground faults) get cleared in a very short time (milliseconds). One of the most common measures of performance of a learning algorithm is accuracy. But, the accuracy measure only draws an effective measure when the classes to be detected are in the same ratio, which is not the case in our domain of study. If 98% of instances are normal and 2% percent are faults, then any "dumb" classifier can achieve 98% accuracy by just labeling each incoming instance as normal. A different evaluation measure has to be used that can evaluate the algorithm regardless of the imbalance in classes.

Kappa Statistics, introduced by Cohen in 1960, is a more appropriate measure to represent the performance of stream classifiers [60]. It normalizes the accuracy by that of the chance predictors which is more credible in our domain of application. The kappa statistic is defined as Equation 9 [60]

$$\kappa = \frac{\rho 0 - \rho C}{1 - \rho C} \qquad \text{Equation 9}$$

where, $\rho 0$ and ρC are prequential accuracy and chance accuracy [20] respectively. If a classifier is always correct, then $\kappa = 1$. If the accuracy coincided with chance classifier then $\kappa = 0$.

A demonstration of calculation of Kappa statistics will make more clear point on its use. Confusion matrix for a hypothetical classifier, which classifies a highly unbalanced classification problem, is shown in Table 8.

Table 8 Confusion matrix for kappa calculation

	Predicted Class			
Real Class		A	B	Total
	A	438	12	450
	B	20	30	50
	Total	458	42	500

Prequential accuracy ($\rho 0$) is $\frac{438+30}{500} = 93.6\%$, which is not a good representation of performance of classifier because 40% of class B is not correctly classified. Kappa statistics gives a better representation as shown below. Chance accuracy is calculated using Equation 10, where, C is confusion matrix, N is number of classes and m is total number of instances.

$$\rho C = \sum_{i=1}^{N} \left(\sum_{j=1}^{N} \frac{C_{ij}}{m}\right) \left(\sum_{j=1}^{N} \frac{C_{ji}}{m}\right) \qquad \text{Equation 10}$$

In our example, $\rho C = 0.8244$

Therefore,

$$\kappa = \frac{0.936 - 0.8244}{1 - 0.8244} = 0.63553 \qquad \text{Equation 11}$$

Kappa statistics punishes algorithms that fail to accurately classify minority class instead of treating both the majority and minority class equally.

Evaluation Time

Evaluation time is time (seconds) required for algorithm to run. Interleaved test then train method of model evaluation does not have clear separation between training and testing phase of an algorithm. A new sample is tested first then model is trained on, so evaluation time consist of both testing time and training time as illustrated Figure 26.

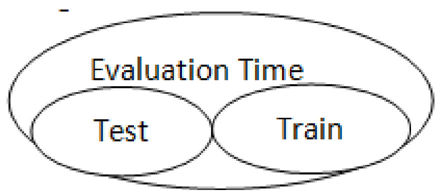

Figure 26 Evaluation time for interleaved-test-then-train

Model Cost (Ram hours)

Ram hours will used as an evaluation measure of the algorithms used for synchrophasor data mining. Every GB of RAM deployed for 1 hour equals one RAM-Hour. Commercial cloud services such as GoGrid which handle huge amount of data charge their customers based on RAM hours for memory usage [92].

Simulation and Results

Experiment I

In this experiment, we focus on testing the ability of stream mining to adapt to changing conditions of power systems. A machine learning algorithm has to constantly update its learned knowledge to stay relevant in predicting behavior of a dynamic system. This is very important feature for an algorithm in order to incorporate dynamic behavior of power systems. In this experiment, we have emulated dynamic behavior by changing loading condition. We generated synchrophasor data with solid three phase faults in various loading conditions. In order to simulate concept drift in the system, Real Power (P) and Reactive Power (Q) are changed at regular intervals. Three phase faults are introduced at regular interval. The simulation is run for about 41 minutes with 74,245 data samples generated.

The magnitudes of parameters shown in table 7 are organized in a row. All magnitudes, frequency, rate of change of frequency and angle difference between phase A Voltages of both PMUs for a timestamp are organized as shown in Table 9. Each of the rows was manually labeled to be normal or fault. A new column named "*class*" was added after manual classification of each row.

Table 9 Organization of samples as training data

Time	PhaseA Mag		Seq1 Mag	..	AngDiff

The data stream mining framework Massive Online Analysis (MOA) [52] is utilized for performing experiments described here. Comma Separated Values (CSV) are converted to an ARFF file format [93], to be fed into MOA. Hoeffding Adaptive Tree (with Naïve Bayes classifier as leaf predictor) and Non-adaptive Hoeffding tree (with Gauss10 numeric estimator) were used to demonstrate the ability of adaptive hoeffding tree to adapt to changing environment. Figure 27 illustrates the performance of both Hoeffding trees (adaptive and non-adaptive) on same set of data based on Kappa Statistics. Non adaptive algorithms always outperformed adaptive algorithm in terms of computation time and memory requirement. This experiment shows that adaptive hoeffding tree achieved better accuracy at expense of runtime and memory. We have used Interleaved-Test-Then-Train approach for model evaluation, so runtime consists of both training and testing, which makes testing time less than reported in these experiments.

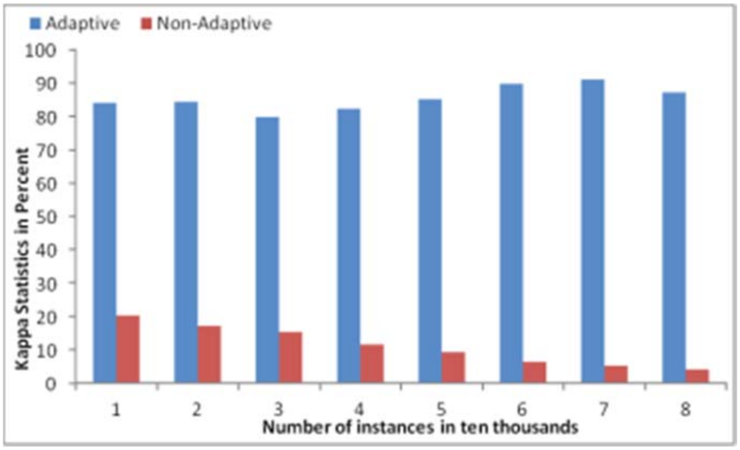

Figure 27 Kappa statistics plot for experiment I

Figure 28 shows a comparison in terms of RAM hours of adaptive and non adaptive hoeffding tree algorithm. Figure 29 shows evaluation time (in seconds) of both algorithms. Non adaptive algorithms always outperformed adaptive algorithm in terms of computation time and memory requirement. This experiment shows that adaptive hoeffding tree achieved better accuracy at expense of runtime and memory. We have used Interleaved-Test-Then-Train approach for model evaluation, so runtime consists of both training and testing, which makes testing time less than reported in these experiments.

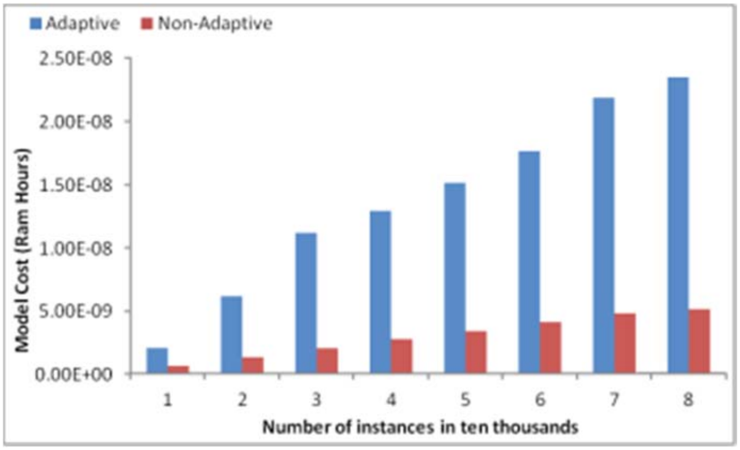

Figure 28 Model cost in ram hours for experiment I

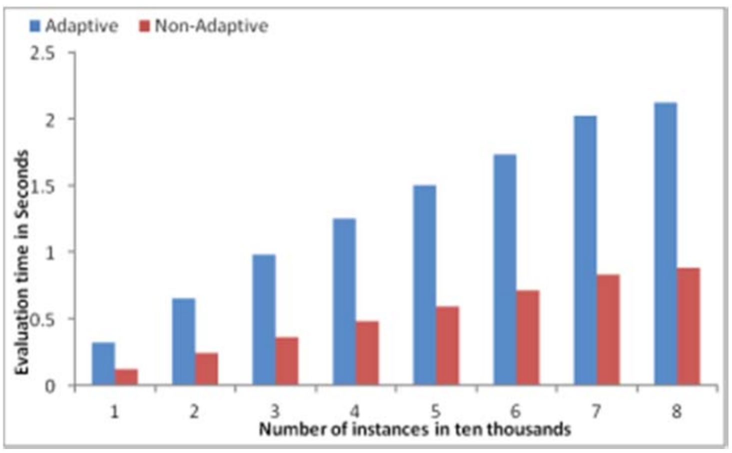

Figure 29 Evaluation time in CPU second for experiment I

Experiment II

In this experiment, 26 Single Line to Ground (SLG) faults were introduced at regular intervals. Phase A to Ground, Phase B to Ground and Phase C to Ground faults are introduced on Bus 1, Bus 4 and Bus 7 each with 100 Ω fault impedance. All other factors such as load (P) and (Q) remained constant throughout the experiment. The simulation was run for about an hour to generate 107,117 data samples. The training data was generated using similar measures as in Experiment I. All kinds of SLG faults were classified into a single category called "FAULT" for binary classification.

In this experiment, loading condition remains constant throughout the experiment. As with experiment I, we tested results of both adaptive and non-adaptive Hoeffding tree. Kappa statistics plot for this experiment is illustrated Figure 30, where accuracy level was fairly constant in mid 90s for adaptive algorithm. As the stream is not evolving in this experiment, non-adaptive Hoeffding tree performed fairly well compared to evolving data stream in Experiment I.

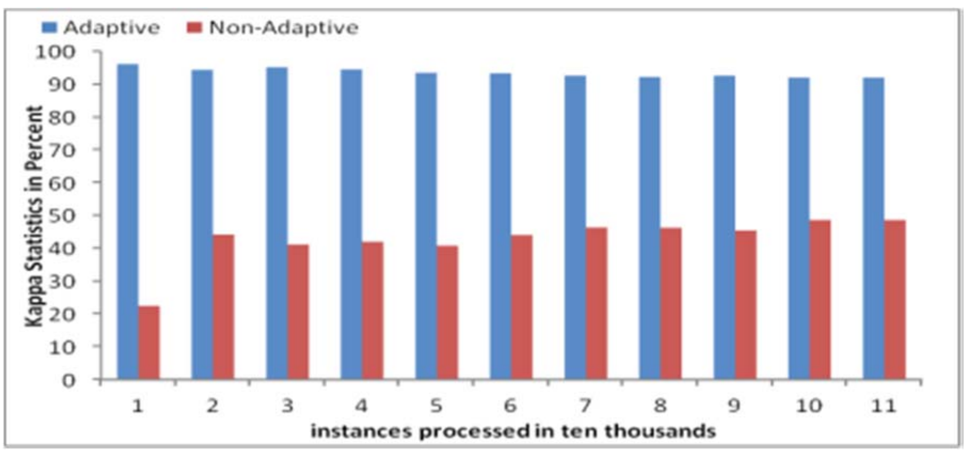

Figure 30 Kappa statistics plot of adaptive and non adaptive hoeffding trees

Although loading condition in this experiment is kept constant, three different types of faults (A-G, B-G and C-G) are categorized in a single class. Unlike batch processing algorithms, stream mining algorithm does not have access to training data at once, so adaptive algorithm seem to be adapting well to different type of faults presented to it as a single class in a stream. The non-adaptive algorithm also seem to be performing better than that in experiment I because variation in data distribution is not as radical as that in experiment I, where loading condition is changing. Also, the performance of non-adaptive algorithm is not as good as the experiment I because different fault types are presented in a stream instead of a batch, so the algorithm could not adapt well to identify different kind of faults categorized as a single class.

Similar to Experiment I, adaptive algorithm was more accurate than non-adaptive algorithm at expense of runtime and memory requirements as shown in Figure 31 and Figure 32.

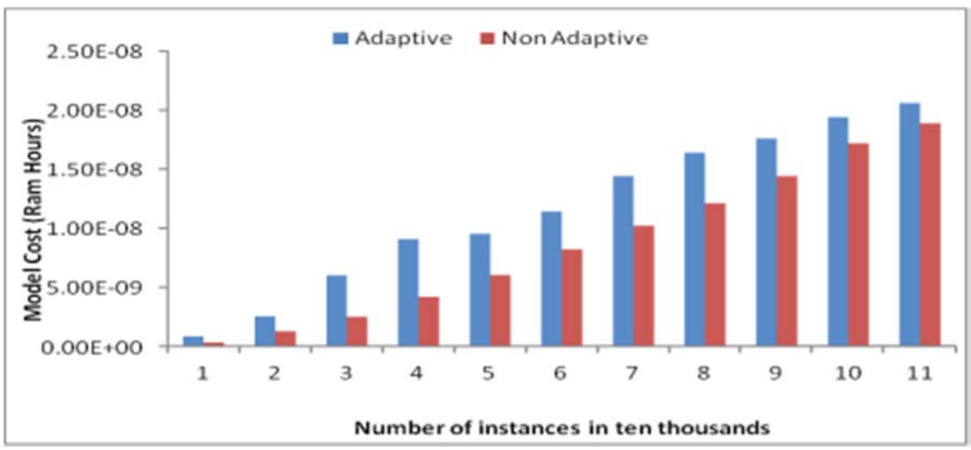

Figure 31 Model cost in ram hours for experiment II

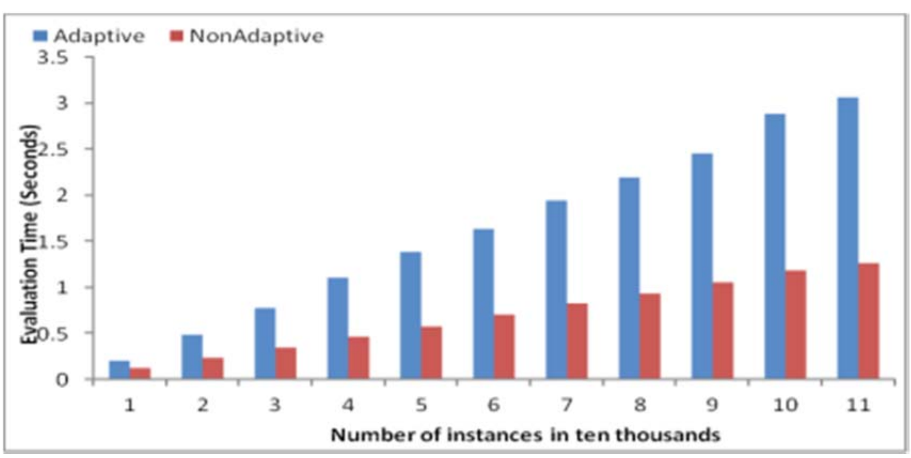

Figure 32 Evaluation time in seconds for experiment II

Experiment III

In this experiment, we fixed the size of hoeffding tree (in bytes) to see effect on accuracy of classification. For the purpose of illustration, we choose non-adaptive hoeffding tree for this experiment. We studied four cases of fixing memory to unbounded memory (memory of host computer), 25K bytes, 50K bytes and 75K bytes. The performances of algorithm for each memory limitation are exactly same, while the unbounded memory performance is better after 140K samples as shown in Figure 33.

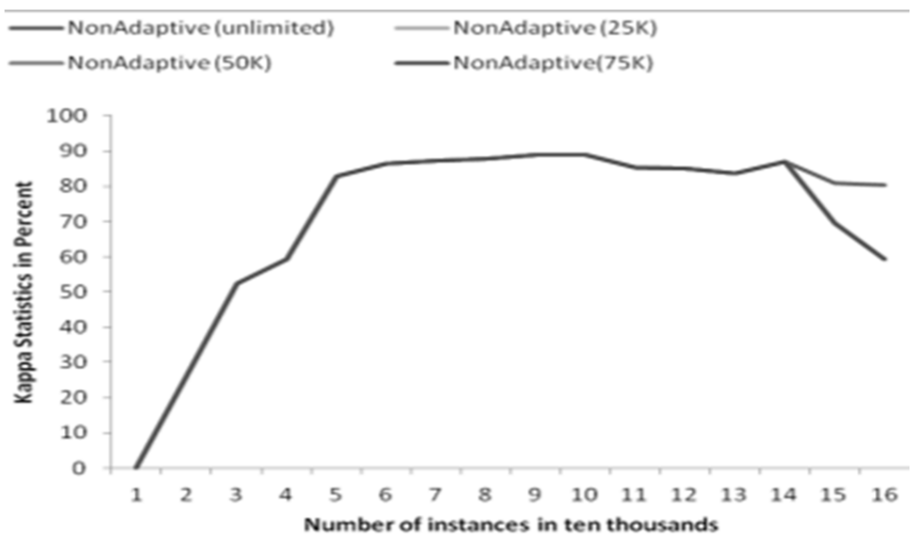

Figure 33 Kappa statistics plot for algorithm with fixed memory

The performance deteriorated when for 25K, 50K, and 75K when the size of tree hit their maximum allocated memory as shown in Figure 34.

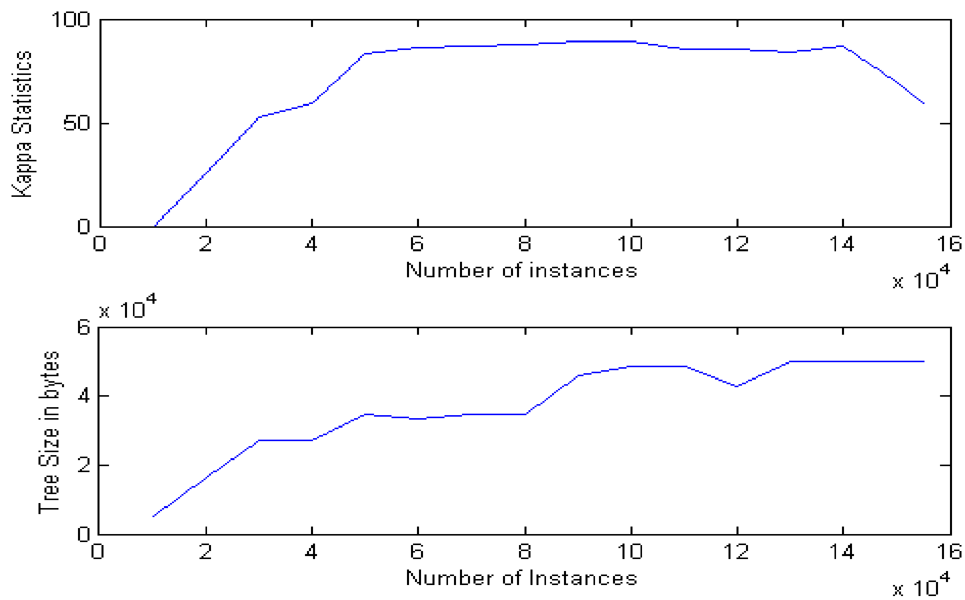

Figure 34 Number of instances vs. tree size vs. kappa statistics for tree size of 50k bytes

This experiment supported our argument that data stream mining algorithms can adapt to lower memory bounds without deteriorating much accuracy.

With number of data samples for a typical power system operation period of 24 hours reaching more than 2 million, the algorithms that process synchrophasor data has to be able to process data in limited memory without affecting much degrading the predictive accuracy. The ability of an algorithm to limit memory use also helps in meeting the latency requirement of real time applications. In this experiment, we have used a small data set to prove that the stream mining algorithm can optimize the tradeoff between memory requirement and accuracy. The impact of this property of stream mining algorithm will be profound when number of data samples is in millions [90].

Experiment IV

In this experiment, we compared the performance of non-adaptive hoeffding tree with traditional decision tree algorithms such as J48 and REPTree available in WEKA [94]. We choose non-adaptive hoeffding tree (unbounded memory) because from Experiments I and II we know that adaptive algorithms are less efficient in terms of memory and runtime, so we did not want to put hoeffding tree in disadvantage for adapting to changes that traditional data mining are not capable of. Figure 35 illustrates performance comparison based on runtime, size of tree and accuracy. The hoeffding tree algorithm significantly outperformed others even when runtime of hoeffding tree contains both testing phase and training phase while run time of J48 and REPTree algorithms is the time to just build model.

Hoeffding tree algorithm was also found better than in terms of efficiency in memory as shown in Figure 35. We used Tree size as a measure of memory resource used by the algorithm because it was the only parameter available for all algorithms under study. The accuracy measure of hoeffding tree is found to be slightly lower than that of J48 and REPTree. It may be because of the fact that hoeffding tree is over pruned version of a tree [87]. If the number of samples is increased then hoeffding tree may even catch up with the accuracy other decision tree algorithms [87]. Nevertheless the performance of hoeffding tree is found to support our proposed method of handing huge amount of synchrophasor data within limited memory resource and latency required by situational awareness applications.

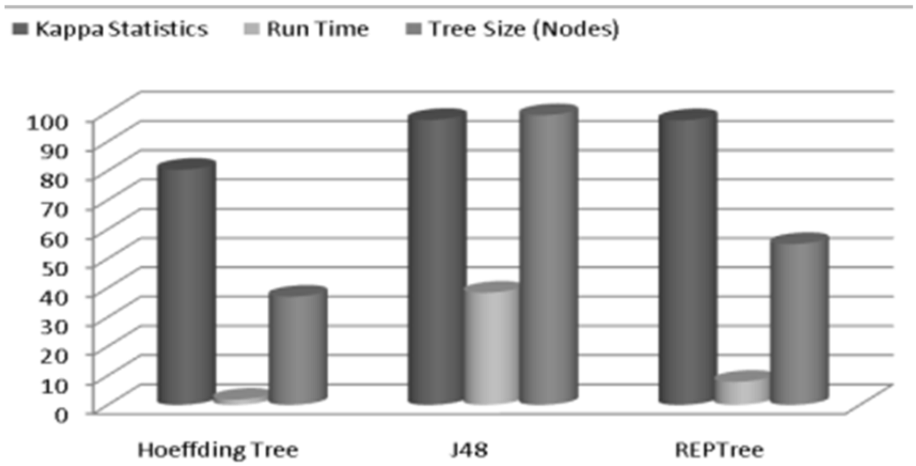

Figure 35 Performance comparison of three algorithms based on runtime, accuracy and memory requirements

Conclusion

Currently, operators do not have a decision support about system wide area status of power systems; they have to rely on local measurements backed by their conscience to intervene an imminent threat to the power system, which may not always be correct. With the deployment of synchrophasors, a huge amount of wide area measurements is available but actionable intelligence inside control room is still lacking. The stream mining algorithm presented in this paper can be used to fill in the information scarcity in situations where operators have to act on system to prevent cascading failures. We presented a new accurate event detection method utilizing massive synchrophasor data while limiting computation and memory requirements. Unlike, conventional machine learning algorithms, data stream mining algorithm is capable of handling stream of data in real time. In addition, the incremental learning method adopted by stream mining algorithms makes it very desirable for application in power system, whose behavior changes very often.

The algorithm presented in this paper can be used to process synchrophasor data within acceptable time while remaining within memory requirements and accuracy for classifying events in power systems. The operators can be alerted quickly about issues to be addressed improving the situational awareness inside control room to improve the reliability of future power systems

CHAPTER V

DIMENSION REDUCTION USING MUTUAL INFORMATION OPTIMIZATION

Background

Synchrophasor technology has been widely regarded as a one of the most important data acquisition technologies for wide area monitoring of power system [95]. Higher data acquisition rate and time synchronization of acquired data have enabled industry to observe a never seen before perspective of dynamic behavior power system. Synchrophasor technology has been termed as MRI scan of power system in comparison to X-Ray for SCADA based monitoring system [96].

Time synchronization enables measurements from multiple PMUs to be temporally aligned as a vector $\{x1,x2,x3\ldots,xn\}$,where 'n' depends upon number of PMUs and number of parameters measured by each PMU. The dimension of synchrophasor vector is destined to rise exponentially with increase in number of deployed PMUs. The dimension of synchrophasor data handled by the Tennessee Valley Authority (TVA) is currently 1850 [62].

The high data acquisition rate and high dimensionality can pose challenge in processing, storing and transferring synchrophasor data efficiently. As machine learning algorithms are finding new applications in supporting decision in power systems [10, 14, 49, 97-99], high dimensionality of synchrophasor data also can impair their performance due to "curse of dimensionality" [100].

Introduction

The infamous northeast blackout of August 14 2003 could have been prevented with adequate level of situational awareness for unforeseen contingencies [101]. Synchrophasor technology provides high speed time-synchronized data for real-time surveillance of power systems. One of the most intriguing areas of research in power systems is to translate massive synchrophasor data into actionable intelligence inside a control room.

The massive amount of synchrophasor data poses a challenge to meet time requirement and accuracy requirement of real-time applications in power systems. Robust pattern recognition and machine learning algorithms are required to identify unforeseen contingencies in real-time to alert stakeholders, so that timely corrective could be taken to prevent cascading failures [102].

A naive way of speeding up data processing is to downsample. But, this method is inappropriate for synchrophasor measurements because it eliminates the essence of using high sampled synchrophasor data. The details of dynamic behavior of power systems are lost. Figure 36 shows plots of 30 samples per second (left) and downsampled version of same signal. It can be observed that much of detail information is lost in downsampled signal. Synchrophasor data should be processed without losing information about dynamic behavior of power systems.

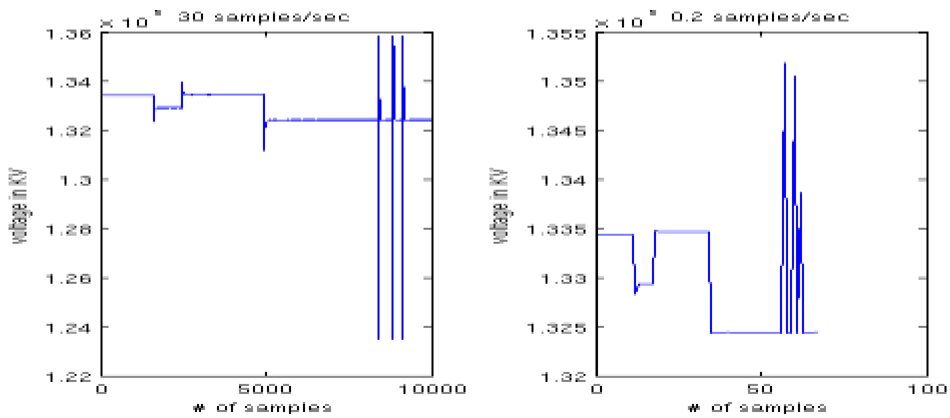

Figure 36 Voltage measurements sampled at 30 samples/second and 0.2 samples/second

In this paper, we propose a feature selection method based on mutual information content of measurements from PMUs. This method will exploit the fact that each synchrophasor measurement does not carry equal/unique information for classification of an event in power systems. The proposed method will reduce redundancies in measurements while increasing relevance of measurements for event detection. A subset of best measurements is selected thus reducing computation complexity without losing much information.

There are several advantages discarding less informative measurements using information theory based criteria:

- Computational cost is reduced without losing information
- Effect of "curse of dimensionality" is reduced
- Interpretability of PMU measurements are increased because the proposed method quantifies importance of measurements and
- Noisy measurements can be discarded thus improving performance of classification algorithms.

Dimension Reduction Techniques

Dimension reduction is a process of reducing the amount of data with minimal loss of information content. With advancements in data collection techniques, most areas of science and engineering are overwhelmed with the amount of data waiting to be analyzed. Dimension reduction is not a new area of study. It has been studied for a long time by researchers in statistics, computer science, machine learning, signal processing etc. There are two major areas of study in dimension reduction.

Feature Extraction

Feature extraction is a method of reducing dimension that extracts relevant and unique information from the data set. In this method, the original signal is mathematically modified and a new set of data with smaller dimensionality is generated as shown in Figure 37.

The importance of a feature may depend on the application. Most Artificial Intelligence (AI) techniques, such as Back Propagation Neural Networks [67], feed forward neural networks [68], Kohonen maps [69], etc are examples of feature extraction. There are several other mathematical feature extraction processes such as Principal Component Analysis (PCA), wavelet methods, Singular Value Decomposition (SVD), etc [70].

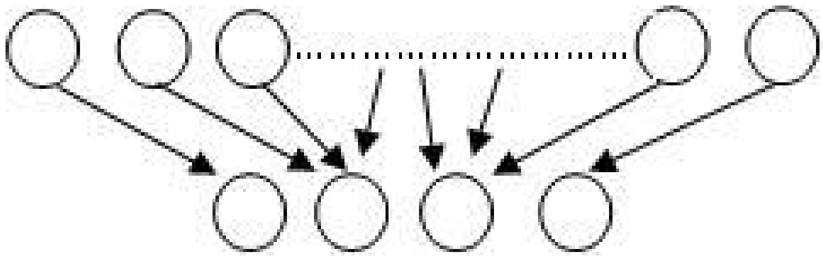

Figure 37 Illustration of dimension reduction process

Feature extraction processes may be either supervised or unsupervised. Supervised methods, such as neural networks, learn from training data. The characteristics of sample data are used by feature extraction methods to generalize the importance of features, while unsupervised mathematical models, such as Principal Component Analysis (PCA), wavelets transform, etc. use mathematical tools to extract energy representing the importance of the features.

Feature Selection

The complete feature set describing a data set carries all of the information of the data. However, a subset of features can often be used to describe certain underlying trends within the data. The feature selection technique chooses a subset of "important features" from the total set of features without altering the original data.

There are several strategies to find an optimal set of "important features". Feature selection techniques can be broadly classified into *feature ranking* and *subset selection* categories. *Feature ranking* methods rank available features based on parameters such as information gain or distance[64]. A subset of highly ranked features is selected as a representative set of the original data set. A decision tree is an example of the feature ranking method. *Subset selection* techniques evaluate a subset of data against a model and modify the model until a satisfactory subset is obtained [64]. Optimization algorithms such as Genetic algorithms fall in this category of subset selection.

The feature selection technique may be supervised or unsupervised. Evolutionary algorithms, such as genetic algorithms, encode domain knowledge as a fitness function [65]. Decision trees require a supervised training approach to select features [103]. Feature selection using clustering algorithms do not require any training [66].

Feature selection methods can be widely classified into two categories: Filters, wrappers and embedded methods.

Filters select a set of best features based on a scoring criterion function *f(i)*. The features yielding largest value of *f(i)* are generally considered best and the features will be sorted in descending order of f(i) for selecting n number of features. This method is independent of specific choice of predictors and is explicitly a pre-processing step [64].

Wrappers select a set of best features based on their usefulness to a given predictor. The final problem is solved with various subsets and subset yielding best results is selected. It is a time consuming method but assures the best result [64].

Embedded methods incorporate feature selection during training phase. Decision Tree is an example of embedded feature selection method.

In this paper, we will use mutual information based feature selection method for selecting most informative features. This method is a filter with optimum mutual information as the criterion function.

Mutual Information Based Feature Selection

Mutual information is the measure of information carried by one random variable about another random variable. It is a measurement of reduction of uncertainty about one variable after knowing value of another variable [104]. If X and Y are two random variables, then mutual information is given by Equation 12.

$$I(X;Y) = H(X) - H(X|Y) \qquad \text{Equation 12}$$

Where, H(X) and H(Y) are the marginal entropies, H (X|Y) and is the conditional entropy.

Entropy is a measure of uncertainty of a random variable. The entropy H(X) of a discrete random variable X is defined by Equation 13.

$$H(X) = -\sum p(x) log\, p(x) \qquad \text{Equation 13}$$

Where, *p(x)* is the probability mass density function of random variable *x*. The log is to the base 2 and entropy is expressed in bits [104].

Conditional entropy is the measure of remaining uncertainty of a random variable when another random variable is known. The formula for conditional entropy is given by Equation 14.

$$H(X|Y) = \sum_{x\epsilon X, y\epsilon Y} p(x,y) \log\left(\frac{p(x)}{p(x,y)}\right) \qquad \text{Equation 14}$$

Where, *p(x,y)* is joint probability distribution function of random variable *x and y*. The log is to the base 2 and conditional entropy is also expressed in bits [104].

If $\{x_1, x_2, x_3, x_4, x_5,.. x_6 ; c\}$ is a synchrophasor vector with c as class label (or state of power system) of the vector, then mutual information *I(C;X)* can be quantitative measure of helpfulness of the measurement to correctly classify state of power systems [105]. The measurements can be simply ranked in descending order to obtain the best set of measurements. A simple ranking method just ensures that the best measurements individually. It does not necessarily ensure the best performance as a group [106, 107]. In other words, "the best m features are not m best features" [64, 107].

The highly informative measurements generally tend to have high correlation. High correlative features do not provide unique information for a classification problem. If redundancy is not eliminated then computational resources are wasted for processing multiple measurements for extracting same information [64, 106]. The method presented in this paper selects the synchrophasor measurements that are most relevant to a state of power system while minimizing information redundancy among selected measurements.

Mutual Information of Synchrophasor Data

The phasor measurements are continuous values while class labels are discrete values. It is computationally expensive to calculate mutual information for continuous variables. The probability density functions are required for calculating mutual information which involves integrations [107-110]. A formulation has been used in [108-110] to calculate mutual information between continuous variable and discrete class labels using parzen window to estimate probability density of continuous variables. We use this criterion to quantify the relevance of synchrophasor measurement to a state of power system.

Maximizing Relevance

If C is classification array of status of power system and X be a synchrophasor measurement array, then the information content can be calculated using Equation 15.

$$I(C;X) = H(C) - H(C|X) \qquad \text{Equation 15}$$

Where H(C) and H(C|X) are marginal entropy and conditional entropy.

As C is a discrete variable, H(C) can be easily calculated, but H(C|X) is computationally expensive to calculate [108-110] using Equation 16 because synchrophasor measurements are continuous variables.

$$H(C|X) = -\int_X p(x) \sum_{c=1}^{N} p(c|x) \log p(c|x)\, dx \qquad \text{Equation 16}$$

where, N is the number of classes

The conditional entropy H(C|X) of synchrophasor measurements can be estimated using parzen window method. By the Bayesian rule (Equation 17),

$$p(c|x) = \frac{p(x|c)p(c)}{p(x)} \qquad \text{Equation 17}$$

If the class has *N* values, we get estimate of conditional pdf $\hat{p}(x|c)$ of each class using parzen window as Equation 18 [109].

$$\hat{p}(x|c) = \frac{1}{n_c}\left(\sum_{i \epsilon I_c} \emptyset(x - x_i, h)\right) \qquad \text{Equation 18}$$

Where $\emptyset$ (.) is the window function and h is the window of width parameter. If $\emptyset$ and h are selected properly, then $\hat{p}$ converges to true probability density [110]. The widow must be normalized to 1 (See Equation 19).

$$\int \emptyset(x, c)dx = 1 \qquad \text{Equation 19}$$

and the width of window should be the function of n such that

$$\lim_{n\to\infty} h(n) = 0 \; and \; \lim_{n\to\infty} nh^d(n) = \infty \qquad \text{Equation 20}$$

If Gaussian window function is used then for one dimensional Gaussian window then

$$\emptyset(x,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x-x_i}{2\sigma^2}} \qquad \text{Equation 21}$$

Where, σ is the standard deviation of the window. We will use σ = 1/log n as suggested in [110].

The conditional entropy with n training samples by replacing integration with summation of the sample points and each sample has the same probability [109], we get

$$\hat{H}(C|X) = -\sum_{j=1}^{n} \frac{1}{n} \sum_{c=1}^{N} \hat{p}(c|x_j) \log \hat{p}(c|x_j) \qquad \text{Equation 22}$$

where, x_j is the jth sample of the training data.

In this paper, we will use $I(C;X)$ to maximize the relevancy of measurements to state of power system. If S be a subset of synchrophasor measurements, then the condition for maximizing total relevance of all measurements in S is given as [107].

$$\max V, \quad V = \frac{1}{|S|} \sum_{X \in S} I(C, X) \qquad \text{Equation 23}$$

Minimizing Redundancy

Equation 23 ensures that the synchrophasor measurements that can individually provide maximum discriminative power for class differentiation. But, it tends to select similar information, increasing redundant information [106, 107]. If the redundancy is minimized in the selected set, then a set of measurements with unique discriminative powers can be obtained, thus optimizing usage of computational resources. If S be a set of synchrophasor measurements and |S| be size of S then

$$\min W, \quad W = \frac{1}{|S|^2} \sum_{i,j \in S} I(i,j) \qquad \text{Equation 24}$$

The estimation of mutual information I (x,y) can be done using the Equation 25 [64]

$$\int_{x} \int_{y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} dx\, dy \qquad \text{Equation 25}$$

where *p(x)* and *p(y)* are the probability densities of x and y and *p(x,y)* is joint density.

In this paper, we are trying to minimize redundancy between two synchrophasor measurements which are both continuous variables. Estimation of p(x) and p(y) are computationally expensive for continuous variables. The measurements selected by Equation 21 tend to have higher correlation [106]. So in order to minimize the redundancy, we use Pearson correlation coefficient as a measure of I(X,Y) [106].

As both positive and negative correlation coefficients are considered as redundancy in information, so we will take absolute value of correlation coefficient. The Equation 25 can be re-expressed as

$$\min W, \quad W = \frac{1}{|S|^2} \sum_{i,j \in S \,\&\, i \neq j} |\Upsilon(i,j)| \qquad \text{Equation 26}$$

Where, γ(i,j) is Pearson correlation coefficient calculated as Equation 27 [111].

$$\gamma(x,y) = \frac{\sum_{i=1}^{N} (x - x_i)(y - y_i)}{\sigma_x \sigma_y} \qquad \text{Equation 27}$$

Maximum Relevance-Minimum Redundancy

In order to optimize the minimum redundancy and maximum relevancy of synchrophasor data, we need to optimize the results provided by Equation 24 and 27 simultaneously. [106, 107] proposed two methods for combining the two conditions as Equation 28.

$$\text{Max } (\alpha{*}V-\beta{*}W) \qquad \text{Equation 28}$$

Where, α and β are the tuning variables, which can be used to weigh importance of V and W depending upon specific problem requirements.

Optimization Problem

Let us consider that λ be the number of features to be selected from available set of N measurements. The value of *V-W* has to be maximized for the selected subset of measurements for optimum information. This essentially converts entire feature selection procedure to an optimization problem of V-W. The exhaustive search of optimum subset may quickly turn into a NP hard problem because the possible combinations of subsets are ${}^{\lambda}C_N = \frac{N!}{\lambda!(N-\lambda)!}$. For example: If 5 features are to be selected from 25 features then possible combinations of features are 53130, making exhaustive search practically infeasible.

As the number of phasor measurement units increases, the number of candidate features increases exponentially for the feature selection method. The exhaustive search may not also be possible because of computational cost of probability density function estimation of continuous variables. It may not be possible to find the optimal solution for the problem, but a near optimal solution can be obtained using several optimization algorithms such as genetic algorithms, random mutation hill climbing, ant colony optimization etc.

Optimization using Random Mutation Hill Climbing

In this paper, we will use Random Mutation Hill Climbing (RMHC) method for information optimization of synchrophasor measurement [112]. The computational cost of genetic algorithm (GA) is significantly larger than that of RHMC because it typically maintains a population size of 100 in contrast to RHMC which has only one individual [113]. It is a waste of computational resources if genetic algorithm is used where less extensive random mutation hill climbing can be used.

The specifications of random mutation hill climbing algorithm be described as following

Chromosome: We will use a chromosome of length λ, where λ is number of measurements to be selected. Each gene can take an index of candidate measurement matrix.

Fitness Function: The objective of the RMHC algorithm is to maximize value of (α V-β W).

The basic random mutation hill climbing algorithm is as follows [112]. A flowchart of the algorithm is illustrated in Figure 38.

- Randomly initialize chromosome. Save the chromosome in a separately as *best_chromosome*.
- If number of iterations is less than 90% of maximum number of iterations, mutate all genes and evaluate its fitness.
- Otherwise mutate one randomly selected gene and evaluate its fitness.
- If fitness of mutated chromosome is better than *best_chromosome*, replace it with new chromosome.
- If maximum number of iterations is reached stop else go to step 2.

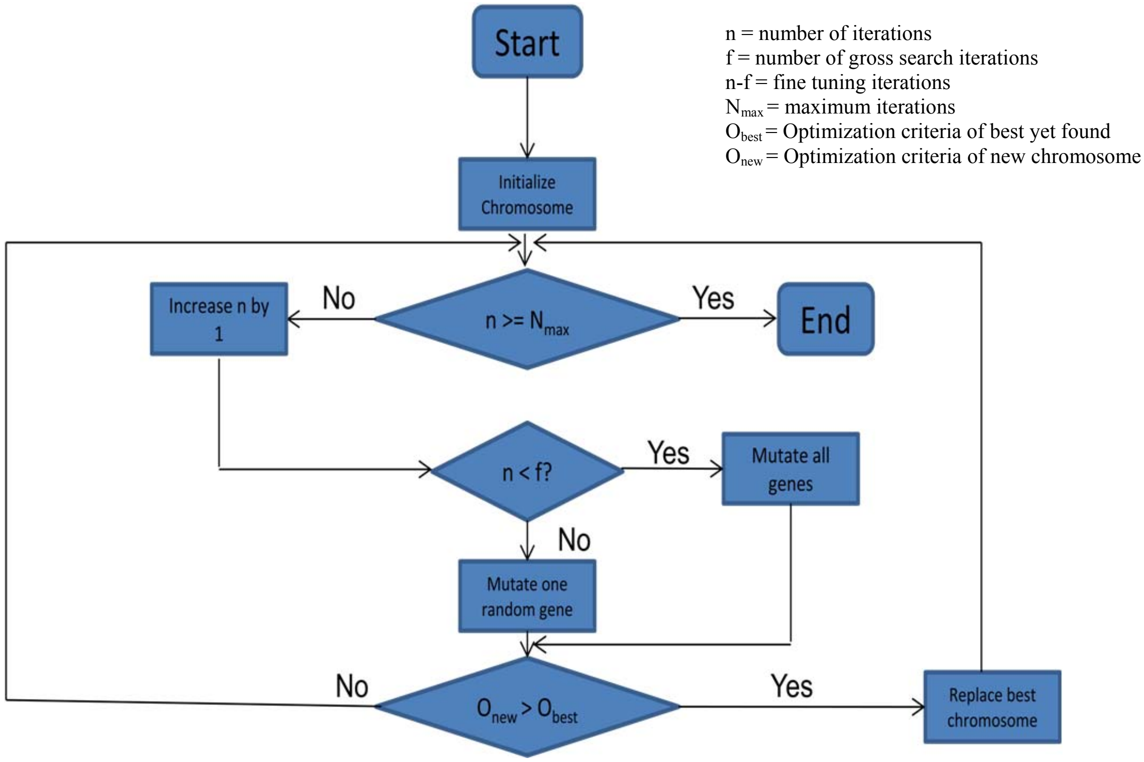

Figure 38 Flowchart of optimization of mutual information

Experimental Settings

Hardware Setup for Data Generation

The data required for the proof of concept proposed in this paper is generated by using a power system modeled in Real Time Digital Simulator (RTDS) with hardware-in-the-loop using two phasor measurements units (PMU), Phasor Data Concentrator (PDC). It can simulate a variety of operating conditions [114]. The simulation test-bed is illustrated in Figure 39. A detailed description of test-bed development and hardware setup can be obtained in [114].

Figure 39	Simulation setup [114]

The test-bed includes two PMUs GE N60 and SEL421. Both PMUs are configured to send data to PDC at 30 samples per second. The PDC concentrates data and sends data to OpenPDC for processing. We have written a custom output adapter on OpenPDC to format data as shown in Figure 40, so that it can be processed by our algorithm.

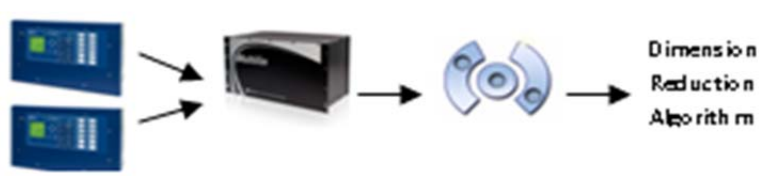

Figure 40	Flow of data from simulation to algorithm

Scenarios

In this paper, we study two scenarios to demonstrate effectiveness of the proposed method of dimension reduction. We focus on being able to predict operating condition of power system with minimum number of synchrophasor measurements instead of all available measurements.

Change of Load

Load change is a normal phenomenon in electrical power system. Load can change with very simple operation as turning on/off of a light. Each change of load changes the operating condition of power system. Change of load may be important for predicting voltage stability, predicting topology changes, fault detection, outage detection etc [115].

Loss of a generator

Loss of a generator is usually not a normal phenomenon. The effect of a generator loss on power system depends upon the contribution of the generator to maintain the load-generation balance. If the generator contributes huge chunk of power being delivered it may push electrical system towards instability while if the generator contributes smaller amount of energy and other generators can pick-up then it will have almost no effect in operation of power system.

Experimental Results

Experiment I

In this experiment, we apply the proposed dimension reduction algorithm on measurement from scenario of different loading conditions in a power system. The initial load on is 84MW, and then the load was changed twice at step of 12 MW. The other two loading conditions were 96MW and 108 MW. The plots of phase A voltage magnitude, phase A current magnitude and phase angle measurements from one the PMUs is illustrated in Figure 41. It can be observed that there are two changes in load where voltage and current have clear fluctuation where as it can hardly be distinguished with just angle.

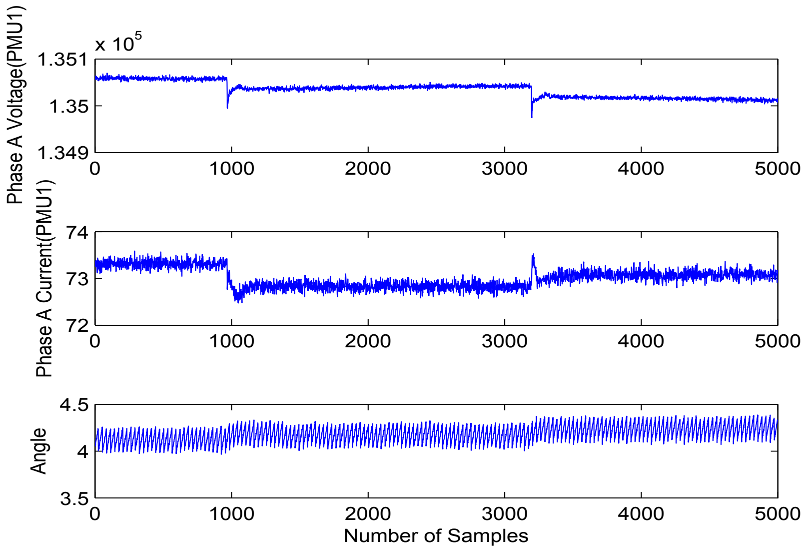

Figure 41 Plot of phase A voltage, phase A current and angle

For this experiment, we have used two Phasor Measurement Units which give 21 different measurements as shown in Table 10. The zero and negative sequence currents and voltage had to be discarded because they were 0 for entire experiment.

Table 10 Synchrophasor measurements used in experiment I

	GE N60	SEL421
All phase Voltages	✓	✓
All phase Currents	✓	✓
Frequency	✓	✓
df/dt	✓	✓
Positive Sequence voltage	✓	✓
Positive Sequence Current	✓	✓
Angle	✓	

Figure 42 illustrates the loading condition discriminative power of each measurement using probability density function of each synchrophasor measurement. It is observed that voltage has the most definitive discriminative power, while current has some overlapping in measurements while most of the measurements in angle are overlapped. This property of synchrophasor measurement is measured by maximum relevancy. The voltage seems to be the most relevant to loading condition while angle seem to be the least relevant.

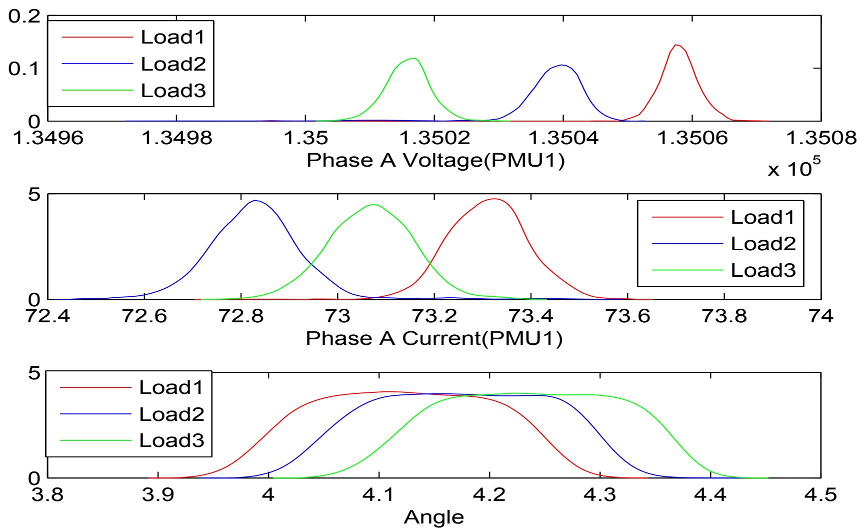

Figure 42 Probability density function of each loading condition

Table 11 shows a sub-matrix of correlation matrix of the synchrophasor measurements. It is observed that the correlation between voltages is very high; the frequency has low correlation with other measurements while correlations between current and other measurements are not consistent. After observing the correlation matrix, we can conclude that a significant correlation exists in synchrophasor measurements and the approach we have taken in this paper to reduce redundancy is relevant to power system domain.

Table 11 Correlation coefficients of synchrophasor measurements

	V	V	I	I	f	V	I	θ
V	1	0.96	0.2	-0.2	0.17	0.85	-0.9	-0.5
V	0.96	1	0.2	-0.2	0.16	0.85	-1	-0.5
I	0.16	0.23	1	0.6	0.01	0.28	-0.3	-0.1
I	-0.2	-0.2	0.6	1	0.05	-0.1	0.1	0.14
f	0.17	0.16	0	0	1	0.12	-0.1	-0
V	0.85	0.85	0.3	-0.1	0.12	1	-0.9	-0.4
I	-0.9	-1	-0.3	0.1	-0.1	-0.9	1	0.48
θ	-0.5	-0.5	-0.1	0.1	-0	-0.4	0.48	1

Now, we use Random Mutation Hill Climbing (RMHC) algorithm to select 5 measurements with maximum non redundant mutual information:

- Length of Chromosome = 5
- $\alpha=\beta=1$
- Fitness Function = max(V-W)
- Iterations =200

The selected features were fed into self organizing feature map using sequential training [116]. Figure 43 shows Self Organizing Map (SOM) created using five most informative measurements selected by proposed algorithm at top while figure at bottom shows SOM created using all synchrophasor measurements available. Visual inspection of the SOMs proves that the clusters are compact and well spaced in top SOM than bottom SOM. More importantly, the structure and relative spatial placement of each cluster is same in both SOM indicating much information is retained in the selected features.

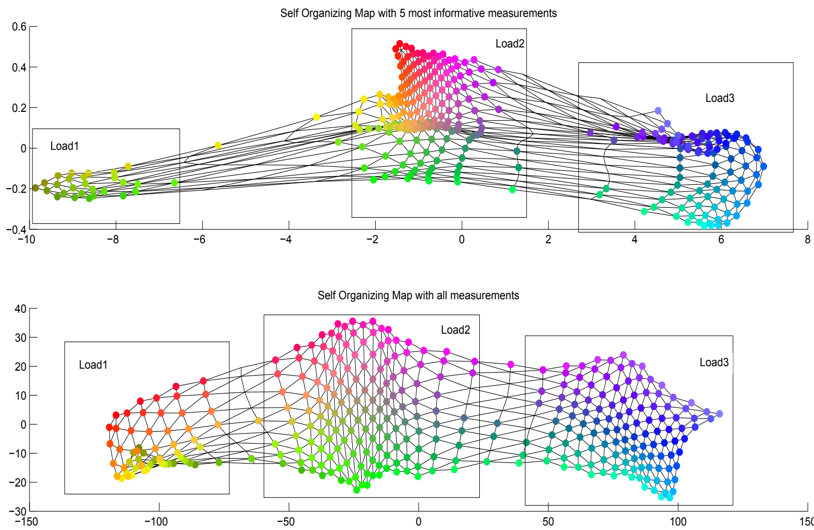

Figure 43 Self organizing map with 5 most informative measurements (top) and with all measurements available (bottom)

In Table 12, we demonstrate performance evaluation of self organizing map algorithm on clustering synchrophasor data. We use CPU time required for creating SOM, size of data and error in classification of supervised self organizing map [116]. We achieved a significant reduction in CPU time and data size also the accuracy of supervised SOM is improved.

Table 12 Performance evaluation of feature selection method

	5 Features	21 Features
CPU Time (Sec)	6.6144	9.7412
Data Size (KB)	66.1	261
Error (%)	2.90	2.96

Experiment II

In this experiment, synchrophasor measurements from a generator loss situation are used to illustrate effectiveness of proposed dimension reduction algorithm. In this experiment, there are 17 synchrophasor measurements from two PMUs, where plot of phase A voltage, phase A current and phase Angle from one of the PMUs is shown in Figure 44. The system is operating at normal condition first then a generator is lost at a point where there is fluctuation in measurements.

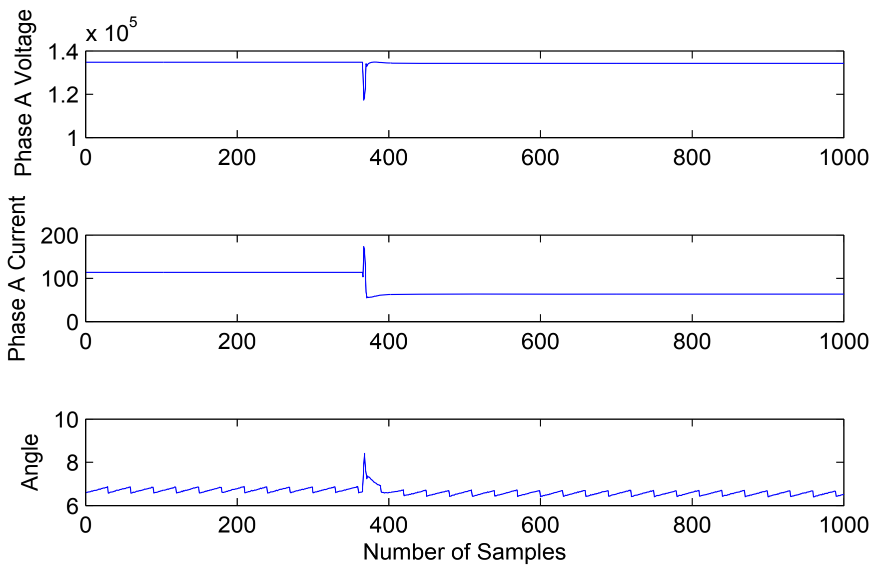

Figure 44 Plot of phase A voltage, phase A current and angle

This experiment is provided a greater fluctuation of measurements than that provided by the experiment I. We considered the transient measurements as outliers for this experiment and filtered out measurement collected within 1 second (30 samples) of the generator loss, to maintain the relevancy of the probability density function to the new operating condition. Table 13 illustrates the properties of data for experiment I, experiment II (before and after transients removed).

Table 13 Properties of synchrophasor data collected

	Exp I	Exp II	Exp II (No Transients)
Phase A Voltage	17.002 KV	528.01KV	232.5037 KV
Phase A Current	0.205 A	24.676 A	24.302 A
Angle	0.0864 Deg	0.156 Deg	0.116 Deg

Now, we use Random Mutation Hill Climbing (RMHC) algorithm to select 3 measurements with maximum non redundant mutual information:

- Length of Chromosome = 3
- $\alpha = \beta = 1$
- Fitness Function = max(V-W)
- Iterations = 200

Figure 45 shows SOM created using five most informative measurements selected by proposed algorithm at top while figure at bottom shows SOM created using all the available synchrophasor measurements. The space and spacing of clusters visually are identical with similar relative position of the clusters. The relative spatial spacing of clusters in both SOM is similar as we observed in experiment I.

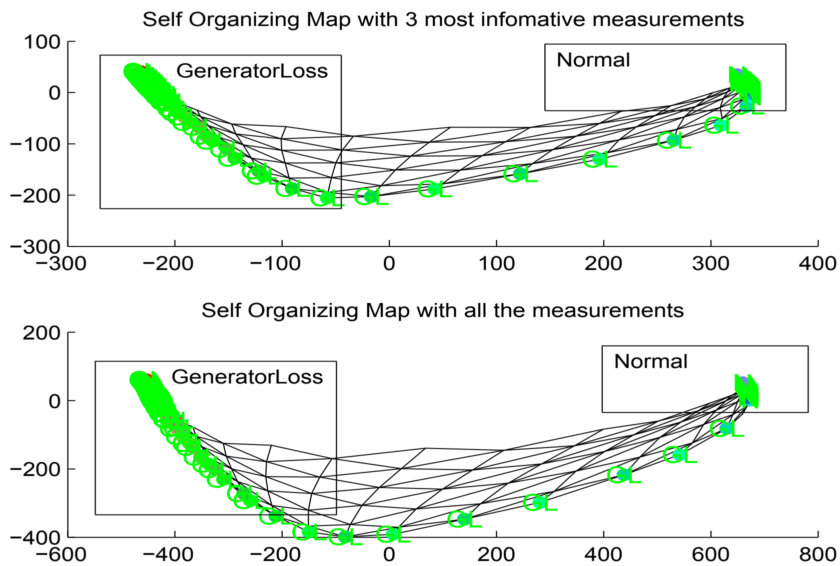

Figure 45 Self organizing map with 3 most informative measurements (top) and with all measurements available (bottom)

Table 14 shows performance measures of the feature reduction technique. CPU time is the time required for creating self organizing map (Figure 45), size of data and percentage of error in classification of supervised self organizing as described in [116]. We achieved significant improvements in all three performance measures using the features selected by our algorithm.

Table 14 Performance evaluation of feature selection method

	3 Features	17 Features
CPU Time (Sec)	1.849670	3.060350
Data Size (KB)	8.06	59.8
Error (%)	0.1024	0.2047

Discussions and Conclusions

In this study, we proposed an algorithm for reducing dimension of synchrophasor measurements based on optimization of underlying information to classify events in power system. The algorithm selects measurements that are more relevant for the classification problem while reducing redundancy in information, thus a set of features carrying unique information.

We performed two experiments to prove the features selected by our algorithm discards irrelevant information to a classification problem and/or redundant to a selected feature. The problem of feature selection is essentially reduced to information optimization of synchrophasor measurements. The computational complexity to exhaustively search the optimum set of features may not practically feasible when the number of available measurements is large, so we used random mutation hill climbing (RMHC) algorithm that can search for a "good enough" set of features.

We used self organizing maps (SOM), an unsupervised learning method to form clusters out of all available measurements and out of measurements selected by our algorithm. The results obtained by measurements selected by our algorithm are evenly matched or even better than the result obtained by using all the measurements. The relative position of clusters are similar in both cases indicating minimal loss of information [117]. The use of the unsupervised learning method to test the information content decouples the supervised learning method that we have used for the feature selection indicating the effectiveness of the algorithm proposed in this chapter.

Limitations and Future Works

The greatest limitation of this method is the computational complexity of estimation of mutual information content of continuous synchrophasor measurements. In future, we can investigate the methods to speed up pdf estimation process with use of modern distributed computation frameworks such as Graphical processors, Hadoop etc.

CHAPTER VI

CONCLUSION AND FUTURE WORKS

Conclusion

Managing and extracting information from large scale data will be a major problem as we move into the future. The massive amount of data collected by a variety of sensors has to be transferred, stored and analyzed efficiently to avoid bottlenecks. The future smart grid will be no exception in generating a massive amount of data. The new sensory equipment (e.g., PMUs, smart meters, frequency disturbance recorders etc) have already started generating a huge amount of data. As the industry is rapidly moving towards achieving a wide area situational awareness system in the near future, rapid data mining algorithms have to be developed to ease certain bottlenecks created by the scale of the data.

As more data is generated, the importance of individual data points will decrease and the importance of analysis of trends of data will increase. The identification of important measurements for event detection will help in discarding "less important" measurements thus, providing savings on processing, storage and transmission cost of the data. In addition, it also adds value on meeting the aggressive latency requirements of the real time situational awareness applications.

In this dissertation, we have proposed various algorithms that contribute towards solving the massive data problem that future smart grid applications will have. The algorithms proposed in this work can be broadly classified into two areas i) Dimension reduction algorithms ii) stream mining algorithms.

Time aligned synchrophasor data enables us to capture a snapshot of the state of the power system at any given point of time. In addition, it also enables comparison of the state of the power system at different geographical locations, thus providing an important forensic tool for post event analysis. The synchrophasor data can be time aligned and treated as a vector of measurements. The dimension of the vector largely depends upon the number of measurements that each PMU can measure. As the number of deployed PMUs increase, the dimension of synchrophasor data increases exponentially. More importantly, each measurement added because of the addition of a PMU will not carry an equal amount of additional information about the state of the power system. If a mechanism of selecting the most informative set of measurements can be devised, then a significant amount of irrelevant and redundant measurements can be removed without compromising much on the information content.

In this dissertation, an online dimension reduction algorithm is proposed that extracts principal components of the signal, so that a predefined fraction of information is retained in the principal components. The number of principal components to be retained is calculated in real time, so that the algorithm can incrementally adapt to abrupt and gradual changes to maintain the information content. The algorithm first decomposes the synchrophasor measurements into principal components (PCs), and then reconstructs the original signal based on the PCs. If the difference between the reconstructed signal and original signal is not within the user defined limit then the algorithm readjusts itself until the condition is met.

Another method proposed in this work is based on a feature selection algorithm that optimizes the mutual information between synchrophasor measurements and state of the power system. It exploits the fact that each measurement does not carry equal amounts of information to identify the state of the power system, so it devises an algorithm that maximizes the relevancy of measurements for classifying status of the power system while it minimizes information redundancy between measurements. An optimization criterion is proposed which can be optimized using various optimization algorithms to find a subset of features that contains unique information.

Synchrophasor data can be viewed as a never ending stream of data flowing to control centers. As long as the communication channels are not disrupted, PMUs continuously send data at a very fast rate. At control centers, the data has to be constantly monitored and operators have to be alerted if any undesired phenomenon is detected. Data at hand has to be processed before a new set of data arrives to prevent the data accumulation, otherwise, ultimately the system will run out of memory. This is also important in satisfying the latency requirement of the real time applications. Data mining algorithms have been studied and applied for emulating the behavior of the power system to meet quick decision requirement of online applications. However, the continuous stream of data challenges the traditional machine learning algorithms designed to address the data scarcity problem. Instead of data scarcity, abundance of data is the major hindrance in applying machine learning algorithms in synchrophasor data. The model grows too large over time to accommodate in memory thus, affecting the recall time.

In order to address the problem, a stream mining algorithm is proposed to process synchrophasor data. Unlike traditional data mining algorithms, stream mining algorithms scan the incoming data only once. Instead of data, it stores the sufficient statistics and it also employs several methods to prevent the model from using too much memory. In addition, the model created by the stream mining algorithm is based on decision trees, so it is easier to prune and re-grow a branch of the tree to incorporate the incremental learning. The incremental learning strategy is very useful in a dynamic system (such as power systems) where the state of the system changes frequently.

Stream mining algorithms can prove a very important tool for supporting decisions of future smart grid because of their efficient memory usage, statistically competitive accuracy, incremental learning strategy and most importantly ability to handle data as soon as it arrives.

In summary, various information mining algorithms are studied in this dissertation to extract information from large amounts of synchrophasor measurements to support real time decision making in situational awareness applications of future smart grid. This dissertation envisions a data processing layer for synchrophasor data utilizing machine learning techniques and information theory to reduce data, but retaining information.

Future Work

In this dissertation, I have mostly focused on utilization of information and a loose comparison of algorithms' performance with traditional data mining algorithms. This work can be extended with benchmarking the performance of the algorithms. Also, this dissertation does not include the architecture specific performance of the algorithms. Massive amount of synchrophasor data processing can be parallelized for information extraction using parallel architectures such as Hadoop

REFERENCES

[1] H. Gharavi and R. Ghafurian, "Smart Grid: The Electric Energy System of the Future," Proceedings of the IEEE, vol. 99, pp. 917-921, Jun 2011.

[2] D. o. Energy, "The Smart Grid: An Introduction," D. o. Energy, Ed., ed.

[3] U. Congress, "Energy independence and security act of 2007," *Washington, DC: Public Law,* pp. 110-140, 2007.

[4] G. W. Arnold, "Challenges and Opportunities in Smart Grid: A Position Article," *Proceedings of the IEEE,* vol. 99, pp. 922-927, 2011.

[5] F. Zakaria, "Steven Chu: 'Swing For The Fences'," in *The Daily Beast*, ed, 2010.

[6] A. Bose, "Power system stability: new opportunities for control," *Stability and control of dynamical systems with applications: a tribute to Anthony N. Michel,* p. 315, 2003.

[7] R. L. King, "Information services for smart grids," 2008, pp. 1-5.

[8] B. Xu and A. Abur, "Observability analysis and measurement placement for systems with PMUs," 2004, pp. 943-946 vol. 2.

[9] NERC. (2010). *Real Time Application of Synchrophasors for Improving Reliability*. Available: https://www.naspi.org/site/StaticPDF/resource/rapir_final_20101017.pdf

[10] L. Wehenkel, "Machine learning approaches to power-system security assessment," *IEEE Expert,* vol. 12, pp. 60-72, 1997.

[11] M. Karthikeyan and V. Malathi, "Wavelet-Support Vector Machine Approach for classification of Power Quality Disturbances," *International Journal of Recent Trends in Engineering,* vol. 1, 2009.

[12] P. G. V. Axelberg, I. Y. H. Gu, and M. H. J. Bollen, "Support vector machine for classification of voltage disturbances," *Power Delivery, IEEE Transactions on,* vol. 22, pp. 1297-1303, 2007.

[13] N. Lin, Z. Jian-guo, D. Zhi-gang, and J. Xiao-ling, "Application of Time Series Forecasting Algorithm via Support Vector Machines to Power System Wide-area Stability Prediction," 2005, pp. 1-6.

[14] S. Y. Y. Z. Q. Qian, "Transient Stability Assessment of Power System Based on Support Vector Machine."

[15] L. Moulin, A. P. A. Da Silva, M. El-Sharkawi, and R. Marks, "Support vector machines for transient stability analysis of large-scale power systems," *Power Systems, IEEE Transactions on,* vol. 19, pp. 818-825, 2004.

[16] D. Q. Zhou, U. Annakkage, and A. D. Rajapakse, "Online monitoring of voltage stability margin using an artificial neural network," *Power Systems, IEEE Transactions on,* vol. 25, pp. 1566-1574, 2010.

[17] D. Niebur and A. J. Germond, "Power system static security assessment using the Kohonen neural network classifier," *Power Systems, IEEE Transactions on,* vol. 7, pp. 865-872, 1992.

[18] P. L. Mao and R. K. Aggarwal, "A novel approach to the classification of the transient phenomena in power transformers using combined wavelet transform and neural network," *Power Delivery, IEEE Transactions on,* vol. 16, pp. 654-660, 2001.

[19] W. M. Lin, C. D. Yang, J. H. Lin, and M. T. Tsay, "A fault classification method by RBF neural network with OLS learning procedure," *Power Delivery, IEEE Transactions on,* vol. 16, pp. 473-477, 2001.

[20] R. Aggarwal, Q. Xuan, R. Dunn, A. Johns, and A. Bennett, "A novel fault classification technique for double-circuit lines based on a combined unsupervised/supervised neural network," *Power Delivery, IEEE Transactions on,* vol. 14, pp. 1250-1256, 1999.

[21] D. Novosel and R. L. King, "Using artificial neural networks for load shedding to alleviate overloaded lines," *Power Delivery, IEEE Transactions on,* vol. 9, pp. 425-433, 1994.

[22] K. Sun, S. Likhate, V. Vittal, V. S. Kolluri, and S. Mandal, "An online dynamic security assessment scheme using phasor measurements and decision trees," *Power Systems, IEEE Transactions on,* vol. 22, pp. 1935-1943, 2007.

[23] R. F. Nuqui, A. G. Phadke, R. Schulz, and N. Bhatt, "Fast on-line voltage security monitoring using synchronized phasor measurements and decision trees," 2001, pp. 1347-1352 vol. 3.

[24] S. Rovnyak, S. Kretsinger, J. Thorp, and D. Brown, "Decision trees for real-time transient stability prediction," *Power Systems, IEEE Transactions on,* vol. 9, pp. 1417-1426, 1994.

[25] E. M. Voumvoulakis and N. D. Hatziargyriou, "Decision trees-aided self-organized maps for corrective dynamic security," *Power Systems, IEEE Transactions on,* vol. 23, pp. 622-630, 2008.

[26] Z. Li and W. Wu, "Phasor Measurements-Aided Decision Trees for Power System Security Assessment," 2009, pp. 358-361.

[27] A. Bifet and R. Kirkby. (2009). *Massive Online Analysis*.

[28] M. Sherwood, D. Hu, and V. Venkatasubramanian, "Real-time detection of angle instability using synchrophasors and action principle," 2007, pp. 1-11.

[29] D. Novosel, V. Madani, B. Bhargava, K. Vu, and J. Cole, "Dawn of the grid synchronization," *Power and Energy Magazine, IEEE,* vol. 6, pp. 49-60, 2008.

[30] E. Martínez, N. Juárez, A. Guzmán, G. Zweigle, and J. León, "Using synchronized phasor angle difference for wide-area protection and control."

[31] J. Quintero, G. Liu, and V. Venkatasubramanian, "An oscillation monitoring system for real-time detection of small-signal instability in large electric power systems," 2007, pp. 1-8.

[32] V. Venkatasubramanian, X. Yue, G. Liu, M. Sherwood, and Q. Zhang, "Wide-area monitoring and control algorithms for large power systems using synchrophasors," 2009, pp. 1-5.

[33] Y. Gong, N. Schulz, and A. Guzman, "Synchrophasor-based real-time voltage stability index," 2006, pp. 1029-1036.

[34] M. Glavic and T. Van Cutsem, "Detecting with PMUs the onset of voltage instability caused by a large disturbance," pp. 1-8.

[35] M. Parashar and M. Jianzhong, "Real Time Dynamics Monitoring System (RTDMS): Phasor Applications for the Control Room," in *System Sciences, 2009. HICSS '09. 42nd Hawaii International Conference on*, 2009, pp. 1-11.

[36] N. J. Nilsson, *Introduction to machine learning*, 1996.

[37] N. Hatziargyriou, "Machine learning applications to power systems," *Machine Learning and Its Applications,* pp. 308-317, 2001.

[38] R. L. King, "Artificial neural networks and computational intelligence," *Computer Applications in Power, IEEE,* vol. 11, pp. 14-16, 18-25, 1998.

[39] R. L. King and D. Novosel, "An architecture for the intelligent detection and alleviation of overloaded transmission lines," *Electric Power Systems Research,* vol. 30, pp. 241-249, 1994.

[40] R. L. King and R. Luck, "Power System Control " in *Wiley encyclopedia of electrical and electronics engineering*, J. G. Webster, Ed., ed: John Wiley, 1999, pp. 17-24.

[41] P. Kundur, N. J. Balu, and M. G. Lauby, *Power system stability and control*, 1994.

[42] N. Dahal, V. Mohan, S. Durbha, A. Srivastava, R. King, N. Younan, and N. Schulz, "Wide area monitoring using common information model and sensor web," 2009, pp. 1-7.

[43] S. Santoso, E. J. Powers, W. M. Grady, and A. C. Parsons, "Power quality disturbance waveform recognition using wavelet-based neural classifier. I. Theoretical foundation," *Power Delivery, IEEE Transactions on,* vol. 15, pp. 222-228, 2000.

[44] H. He and J. A. Starzyk, "A self-organizing learning array system for power quality classification based on wavelet transform," *Power Delivery, IEEE Transactions on,* vol. 21, pp. 286-295, 2006.

[45] C. Chunling, X. Tongyu, P. Zailin, and Y. Ye, "Power quality disturbances classification based on multi-class classification SVM," pp. 290-294.

[46] S. Santoso, E. J. Powers, W. M. Grady, and A. C. Parsons, "Power quality disturbance waveform recognition using wavelet-based neural classifier. II. Application," *Power Delivery, IEEE Transactions on,* vol. 15, pp. 229-235, 2000.

[47] S. Santoso, J. Lamoree, W. M. Grady, E. J. Powers, and S. C. Bhatt, "A scalable PQ event identification system," *Power Delivery, IEEE Transactions on,* vol. 15, pp. 738-743, 2000.

[48] B. H. Chowdhury and K. Wang, "Fault classification using Kohonen feature mapping," 1996, pp. 194-198.

[49] K. Silva, B. Souza, and N. Brito, "Fault detection and classification in transmission lines based on wavelet transform and ANN," *Power Delivery, IEEE Transactions on,* vol. 21, pp. 2058-2063, 2006.

[50] C. W. Liu, M. C. Su, S. S. Tsay, and Y. J. Wang, "Application of a novel fuzzy neural network to real-time transient stability swings prediction based on synchronized phasor measurements," *Power Systems, IEEE Transactions on,* vol. 14, pp. 685-692, 1999.

[51] H. Mori, Y. Tamaru, and S. Tsuzuki, "An artificial neural-net based technique for power system dynamic stability with the Kohonen model," *Power Systems, IEEE Transactions on,* vol. 7, pp. 856-864, 1992.

[52] I. Kamwa, S. Samantaray, and G. Joos, "Catastrophe Predictors From Ensemble Decision-Tree Learning of Wide-Area Severity Indices," *Smart Grid, IEEE Transactions on,* vol. 1, pp. 144-158, 2010.

[53] R. Diao, V. Vittal, and N. Logic, "Design of a Real-Time Security Assessment Tool for Situational Awareness Enhancement in Modern Power Systems," *Power Systems, IEEE Transactions on,* vol. 25, pp. 957-965, 2010.

[54] D. J. Sobajic and Y. H. Pao, "Artificial neural-net based dynamic security assessment for electric power systems," *Power Systems, IEEE Transactions on,* vol. 4, pp. 220-228, 1989.

[55] F. Hashiesh, H. Mostafa, I. Helal, and M. Mansour, "A wide area synchrophasor based ANN transient stability predictor for the Egyptian Power System," pp. 1-7.

[56] A. Khatib, R. Nuqui, M. Ingram, and A. Phadke, "Real-time estimation of security from voltage collapse using synchronized phasor measurements," 2004, pp. 582-588 Vol. 1.

[57] X. Chu and Y. Liu, "Real-time transient stability prediction using incremental learning algorithm," pp. 1565-1569 Vol. 2.

[58] Y. Sun, M. S. Kamel, and Y. Wang, "Boosting for learning multiple classes with imbalanced class distribution," 2006, pp. 592-602.

[59] Q. Gu, Z. Cai, L. Zhu, and B. Huang, "Data mining on imbalanced data sets," 2008, pp. 1020-1024.

[60] A. Bifet and E. Frank, "Sentiment knowledge discovery in Twitter streaming data," 2010, pp. 1-15.

[61] A. J. Viera and J. M. Garrett, "Understanding interobserver agreement: the kappa statistic," *Fam Med,* vol. 37, pp. 360-363, 2005.

[62] *OpenPDC.* Available: http://openpdc.codeplex.com/

[63] S. Papadimitriou, J. Sun, and C. Faloutsos, "Streaming pattern discovery in multiple time-series," 2005, pp. 697-708.

[64] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *The Journal of Machine Learning Research,* vol. 3, pp. 1157-1182, 2003.

[65] D. E. Goldberg, *Genetic algorithms in search, optimization, and machine learning*: Addison-wesley, 1989.

[66] J. G. Dy and C. E. Brodley, "Feature selection for unsupervised learning," *The Journal of Machine Learning Research,* vol. 5, pp. 845-889, 2004.

[67] R. Hecht-Nielsen, "Theory of the backpropagation neural network," 1989, pp. 593-605 vol. 1.

[68] *Feedforward Neural Networks*. Available: http://reference.wolfram.com/applications/neuralnetworks/NeuralNetworkTheory/2.5.1.html

[69] T. Kohonen, "The self-organizing map," *Proceedings of the IEEE,* vol. 78, pp. 1464-1480, 1990.

[70] I. K. Fodor, "A survey of dimension reduction techniques," *Livermore, CA: US DOE Office of Scientific and Technical Information,* vol. 18, 2002.

[71] M. Ye, X. Li, and M. E. Orlowska, "Supervised dimensionality reduction on streaming data," 2007, pp. 674-678.

[72] V. Tomenko, "Online dimensionality reduction using competitive learning and Radial Basis Function network," *Neural Networks,* 2011.

[73] H. Gharavi and R. Ghafurian, "Smart Grid: The Electric Energy System of the Future [Scanning the Issue]," *Proceedings of the IEEE,* vol. 99, pp. 917-921, 2011.

[74] S. Litos Strategic Communication under contract No. DE-AC26-04NT41817, "The SMART GRID: an introduction," T. U. S. D. o. Energy, Ed., ed.

[75] R. Avila-Rosales, M. J. Rice, J. Giri, L. Beard, and F. Galvan, "Recent experience with a hybrid SCADA/PMU on-line state estimator," in *Power & Energy Society General Meeting, 2009. PES '09. IEEE*, 2009, pp. 1-8.

[76] T. N. S. Group, "Technical Analysis of the August 14, 2003, Blackout: what Happened, Why, and What Did We Learn? Report to the NERC Board of Trustees ".

[77] R. L. King, "Information services for smart grids," in *Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE*, 2008, pp. 1-5.

[78] D. Novosel, "Keynote: A comprehensive defense plan against wide area disturbances," 2005, pp. 8-8/36.

[79] L. Chih-Wen, S. Mu-chun, T. Shuenn-Shing, and W. Yi-Jen, "Application of a novel fuzzy neural network to real-time transient stability swings prediction based on synchronized phasor measurements," *Power Systems, IEEE Transactions on,* vol. 14, pp. 685-692, 1999.

[80] R. F. Nuqui, A. G. Phadke, R. P. Schulz, and N. Bhatt, "Fast on-line voltage security monitoring using synchronized phasor measurements and decision trees," in *Power Engineering Society Winter Meeting, 2001. IEEE*, 2001, pp. 1347-1352 vol.3.

[81] S. Kai, S. Likhate, V. Vittal, V. S. Kolluri, and S. Mandal, "An Online Dynamic Security Assessment Scheme Using Phasor Measurements and Decision Trees," *Power Systems, IEEE Transactions on,* vol. 22, pp. 1935-1943, 2007.

[82] I. Kamwa, S. R. Samantaray, and G. Joos, "Catastrophe Predictors From Ensemble Decision-Tree Learning of Wide-Area Severity Indices," *Smart Grid, IEEE Transactions on,* vol. 1, pp. 144-158, 2010.

[83] D. Ruisheng, V. Vittal, and N. Logic, "Design of a Real-Time Security Assessment Tool for Situational Awareness Enhancement in Modern Power Systems," *Power Systems, IEEE Transactions on,* vol. 25, pp. 957-965, 2010.

[84] N. C. N. Chu, A. Williams, R. Alhajj, and K. Barker, "Data stream mining architecture for network intrusion detection," in *Information Reuse and Integration, 2004. IRI 2004. Proceedings of the 2004 IEEE International Conference on*, 2004, pp. 363-368.

[85] X. Liu, "Data Stream Mining In Financial Securities Databases Doctoral Dissertation submitted to City University Of Hong Kong, ," November 2007.

[86] J. W. Seifert, "Data mining and homeland security: An overview," 2007.

[87] P. Domingos and G. Hulten, "Mining high-speed data streams," 2000, pp. 71-80.

[88] A. Bifet and R. Gavalda, "Learning from time-changing data with adaptive windowing," 2007, pp. 443-448.

[89] A. Srivastava, K. G. Ravikumar, and G. Zweigle, "Wide-area monitoring and control using the real time digital simulator and a synchrophasor vector processor," *European Transactions on Electrical Power,* 2010.

[90] Data Stream Mining, A Practical Approach. Available: http://moa.cs.waikato.ac.nz/wp-content/uploads/2010/05/StreamMining.pdf

[91] E. Martínez, N. Juárez, A. Guzmán, G. Zweigle, and J. León, "Using synchronized phasor angle difference for wide-area protection and control," *WPRC, Spokane, WA, USA,* 2006.

[92] A. Bifet, G. Holmes, B. Pfahringer, and E. Frank, "Fast perceptron decision tree learning from evolving data streams," *Advances in Knowledge Discovery and Data Mining,* pp. 299-310, 2010.

[93] F. F. Wu, K. Moslehi, and A. Bose, "Power system control centers: past, present, and future," *Proceedings of the IEEE,* vol. 93, pp. 1890-1908, 2005.

[94] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD Explorations Newsletter,* vol. 11, pp. 10-18, 2009.

[95] A. G. Phadke, "Synchronized phasor measurements-a historical overview," 2002, pp. 476-479 vol. 1.

[96] J. Giri, D. Sun, and R. Avila-Rosales, "Wanted: A more intelligent grid," *IEEE Power and Energy Magazine,* vol. 7, pp. 34-40, 2009.

[97] S. Madan, W. K. Son, and K. Bollinger, "Applications of data mining for power systems," 1997, pp. 403-406 vol. 2.

[98] L. A. Wehenkel, *Automatic learning techniques in power systems*: Kluwer Academic Publishers, 1998.

[99] H. Yoshida, K. Kawata, Y. Fukuyama, S. Takayama, and Y. Nakanishi, "A particle swarm optimization for reactive power and voltage control considering voltage security assessment," *Power Systems, IEEE Transactions on,* vol. 15, pp. 1232-1239, 2000.

[100] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *Information Theory, IEEE Transactions on,* vol. 14, pp. 55-63, 1968.

[101] G. Andersson, P. Donalek, R. Farmer, N. Hatziargyriou, I. Kamwa, P. Kundur, N. Martins, J. Paserba, P. Pourbeik, and J. Sanchez-Gasca, "Causes of the 2003 major grid blackouts in North America and Europe, and recommended means to improve system dynamic performance," *Power Systems, IEEE Transactions on,* vol. 20, pp. 1922-1928, 2005.

[102] A. Chakrabortty. (2011, Handling the Data Explosion in Tomorrow's Power Systems. *IEEE Smart Grid Newsletter*. Available: http://smartgrid.ieee.org/newsletter/september-2011/159-handling-the-data-explosion-in-tomorrows-power-systems

[103] J. R. Quinlan, "Induction of decision trees," *Machine learning,* vol. 1, pp. 81-106, 1986.

[104] J. A. T. Thomas M. Cover, "Entropy, Relative Entropyand Mutual Information," in *Elements of Information Theory*, ed: John Wiley & Sons, Inc., 1991.

[105] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *Neural Networks, IEEE Transactions on,* vol. 5, pp. 537-550, 1994.

[106] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," 2003, pp. 523-528.

[107] 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,* vol. 27, pp. 1226-1238, 2005.

[108] K. Torkkola, "Feature extraction by non parametric mutual information maximization," *The Journal of Machine Learning Research,* vol. 3, pp. 1415-1438, 2003.

[109] N. Kwak and C. H. Choi, "Input feature selection by mutual information based on Parzen window," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 24, pp. 1667-1671, 2002.

[110] J. Biesiada, W. Duch, A. Kachel, K. Maczka, and S. Palucha, "Feature ranking methods based on information entropy with Parzen windows," 2005, p. 1.

[111] S. Wright, "Correlation and causation," *Journal of agricultural research,* vol. 20, pp. 557-585, 1921.

[112] D. B. Skalak, "Prototype and feature selection by sampling and random mutation hill climbing algorithms," 1994, pp. 293-301.

[113] D. J. Stracuzzi, "Randomized Feature Selection," in *Computational methods of feature selection*, H. Liu, Motoda, H., Ed., ed: Chapman & Hall, 2008.

[114] U. Adhikari, T. H. Morris, N. Dahal, S. Pan, R. L. King, N. H. Younan, and V. Madani, "Development of Power System Test Bed for Data Mining of Synchrophasors Data, Cyber-Attack and Relay Testing in RTDS," presented at the 2012 IEEE PES General Meeting, San Diego, CA, 2012.

[115] P. Kessel and H. Glavitsch, "Estimating the Voltage Stability of a Power System," *Power Delivery, IEEE Transactions on,* vol. 1, pp. 346-354, 1986.

[116] J. Vesanto, J. Himberg, E. Alhoniemi, and J. Parhankangas, "Self-organizing map in Matlab: the SOM toolbox," 1999, pp. 35-40.

[117] G. M. Foody, "Applications of the self-organising feature map neural network in community data analysis," *Ecological Modelling,* vol. 120, pp. 97-107, 1999.