Expert Systems with Applications 42 (2015) 6853–6863

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Event stream processing for improved situational awareness in the smart grid

N. Dahal [a,1], O. Abuomar [b,*], R. King [a,b], V. Madani [c]

[a] Department of Electrical and Computer Engineering, Mississippi State University, MS 39762, USA
[b] Center for Advanced Vehicular Systems (CAVS), Mississippi State University, MS 39762, USA
[c] Pacific Gas and Electric (PG&E) Company, San Francisco, CA 94105, USA

A R T I C L E I N F O

Article history:
Available online 9 May 2015

Keywords:
Data mining
Situational awareness
Stream processing
Synchrophasor
Wide area monitoring

A B S T R A C T

Deployment of Phasor Measurement Units (PMU) in the United States transmission grid has brought a new data stream to be processed and an opportunity to improve situational awareness on the grid. This new data stream offers opportunity for a faster detection and response algorithm to minimize wide spread outages. High rate of data collection of PMU systems has also brought a challenge on how to extract information from fast moving PMU data stream in real time to improve situational awareness inside a control room. Despite the fact that mathematical and probabilistic methods are the most accurate methods of stability analysis, online decision making algorithms cannot afford the latency brought by those methods. Traditional batch processing Artificial Intelligence (AI) techniques have been extensively studied as potential replacements for these approaches, however conventional AI techniques do not deal with continuous streams of fast moving phasor data. This paper presented a novel application of the stream mining algorithms for synchrophasor data to meet quick decision making requirement of future situational awareness applications in power systems. To prove that the proposed methods are efficient and capable of handling huge amounts of data with reasonable accuracy and within limited resources of memory and computational power, four different experiments with different conditions (changing/unchanging the load conditions of Real Power and Reactive Power, fixing the size of memory, and comparing the performance of non-adaptive Hoeffding tree with traditional decision tree algorithms) were conducted. The algorithms discussed in this paper support decisions inside the control rooms helping stakeholders make informed decisions to improve reliability of the future smart grid.

1. Introduction

The operations of electrical power system largely depend upon the ability of utilities and reliability coordinators to correctly forecast demand, generation and possible contingencies. The forecasting mechanism is currently accurate enough to ensure 99.97% reliability of current electrical system. But, the stakes in power system are so high that 150 billion dollars in losses are blamed on outages and interruptions, which accounts for just 0.03% of time in grid operations (S. Litos Strategic Communication under contract No. DE-AC26-04NT41817).

The current electrical system, which depends upon the accuracy in forecasting of load, generation and contingency may not be able to keep operation highly reliable under uncertainties to be introduced in future electric grid in the form of renewable sources integration, de-regularized market driven industry, distributed generation etc. The electrical power system has to adapt to handle the unpredictability of components in the future grid while improving the reliability of grid operation.

The reliability of future grid hinges on the ability of operators to detect and intervene against anomalous behavior of the power systems to prevent cascading failures. A constant surveillance of grid health parameters such as frequency, voltage and phase angles, and quick alerts to operators based on trends in the monitored parameters will be instrumental in grid operation, in addition to the forecasting mechanism.

Generally, power systems are designed to withstand a host of pre-determined contingencies with automatic protection and control algorithms. In case of a rare combination of contingencies

---

* Corresponding author at: Center for Advanced Vehicular Systems (CAVS), 200 Research Blvd., Mississippi State University, MS 39762, USA. Tel.: +1 662 325 5576; fax: +1 662 325 5433.

E-mail addresses: nischal.dahal@alstom.com (N. Dahal), abuomar@cavs.msstate.edu (O. Abuomar), rking@cavs.msstate.edu (R. King), VXM6@pge.com (V. Madani).

[1] Present address: ALSTOM Grid, 10865 Willows Rd NE, Redmond, WA 98052, USA.

http://dx.doi.org/10.1016/j.eswa.2015.05.003
0957-4174/

automatic protection systems can fail resulting in a wide spread outage affecting millions of customers (T.N.S. Group, 2003). In transient events, reaction time is at most 100 milliseconds and therefore, automatic control equipment takes over the decision making process with no human intervention in the loop. For long term stability, operators usually have enough time to run simulations and refer to a knowledge base, to make informed decisions (King, 2008). However, there are times in between those two extremes in which human intervention is required but this often happens when there is insufficient information available to support their decision. The wide area situational awareness applications (where a predominant concern in application operation, based on a descriptive view of decision making and "the perception of the applications' elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future" can be defined (Endsley, 1995)) aim to fill the information void in that problem domain, where human intervention is required within short response time to prevent cascading failures (King, 2008).

Synchrophasor technology, supported by high speed communication links, has been widely regarded as one of most important tools for real time wide area monitoring and situational awareness in power system. The high sample rate synchrophasor of technology has enabled the industry to capture details of the dynamic behavior of the electrical system. As an analogy, synchrophasor technology is to SCADA based monitoring what an MRI scan is to X-ray (Giri, Sun, & Avila-Rosales, 2009). The investigation of August 14, 2003 blackout pointed out that if phasor data had been monitored, the blackout could have been prevented. A number of clues surrounding the blackout are missed due to the lack of situational awareness infrastructure (NERC, 2010).

With the deployment of PMUs, the industry now has the capability of monitoring grid health parameters in real time. However, if underlying information in the 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 analysis, contingency analysis, state estimation, etc. are used in power systems. The latency suffered by mathematical calculations makes them unfit for real time situational awareness applications. Machine-learning algorithms, such as, Artificial Neural Networks (ANN) (Chih-Wen, Mu-chun, Shuenn-Shing, & Yi-Jen, 1999) and decision trees (DT) (Kai, Likhate, Vittal, Kolluri, & Mandal, 2007; Kamwa, Samantaray, & Joos, 2010; Nuqui, Phadke, Schulz, & Bhatt, 2001; Rovnyak, Kretsinger, Thorp, & Brown, 1994; Ruisheng, Vittal, & Logic, 2010), 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 available "stored" data samples. 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 (Kai et al., 2007; Kamwa et al., 2010; Nuqui et al., 2001; Rovnyak et al., 1994; Ruisheng et al., 2010). 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 makes them an ideal choice for power systems. However, the number of samples from a PMU doubles after the deployment of every PMU in the application. For example, in a 24 hour period a single PMU produces 2,592,000 samples for a single parameter at 30 samples per second but there will be 5,184,000 samples in a 24-hour period if the system contains two PMUs. With the limited computational time and memory available in computer resources, this can limit the size of the decision trees built using traditional machine learning algorithms. Therefore, it may be hard to accommodate a huge decision tree in a limited computer memory without losing information.

One of the easiest methods to handle huge amount of data is to downsample to an appropriate level. This approach is not appropriate for synchrophasor data because the dynamic behavior of the power system is not properly represented in downsampled data, which may even undermine the advantage of using high speed synchrophasor data. Fig. 1 illustrates the disadvantage of downsampling phasor data. Left half of the figure is PMU data at 30 samples per second (typical PMU data rate), while the right half is downsampled version of the same data at 0.2 samples per second (typical SCADA data rate). The details captured by PMU are lost when it is downsampled. 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 the PMUs is important to portray the dynamic behavior of the power systems and to detect events.

A new method known as data stream mining can extract information from high speed data streams facilitating decision-making within constraints of limited resources and time (Domingos & Hulten, 2000). A decision tree is built from data stream in limited memory using Hoeffding bound to guarantee that the result obtained is as good as that of conventional batch processing algorithms (Domingos & Hulten, 2000).

In this paper, we present an application of the stream mining algorithm for synchrophasor data to meet quick decision making requirement of future situational awareness applications in power systems in a limited memory and with limited computational power.

2. Related work

Data stream mining has recently been employed as a technique to analyze and study the data streams of different power systems for knowledge discovery and for addressing some of the related challenges. For ensuring the security of the household electricity appliances, Ma, Fang, Yuan, & Wang, 2014 designed a power security system based on stream data mining, which is mainly composed of the intelligent electric outlet, the coordinator and the server. In the system, the ZigBee module as a connector between the traditional power grid and the coordinator was used and the intelligent electric outlet to shut or open the power was implemented as well. In addition, considering a power grid application in which thousands of sensors are deployed on the power grid network to continuously collect streams of digital data for real-time situational awareness and system management, Omitaomu, 2014 proposed a new data processing and transformation approach, based on the concept of control charts, for representing sequence of data streams from sensors. In addition, an application of the proposed approach for enhancing data mining tasks such as clustering using real-world power grid data streams is presented. Chen, Yang, Xu, and Yuan (2014) designed a composite classifier as a combination of a cache-based classifier (CBC) and a main tree classifier (MTC). This classifier can handle high-speed data streams collected from power grid units and is used as a decision support system that converts the data streams to operational intelligence. Reinhardt and Koessler (2014) presented a concept for the efficient local storage and processing of power consumption data called PowerSAX. Instead of operating on raw sensor data streams readings, PowerSAX converts consumption data into their symbolic representations and thus mitigates their storage requirement. Yang, Chen, Yuan, and Lianhang (2014) established a decision management system in a form of rules and patterns to forecast the power load around the smart grid system. The findings are presented in an easily understood way that can be accepted by both human beings and machines. Bank, Omitaomu, Fernandez, and Liu (2009), presented a method for visualizing high-speed data