INL/RPT-23-76029-Revision-0

Power System Waveform Datasets for Machine Learning

January 2024

Changing the World's Energy Future

Christopher Roger Sticht, Stephen Arthur Bukowski

Idaho National Laboratory

INL is a U.S. Department of Energy National Laboratory operated by Battelle Energy Alliance, LLC

INL/RPT-23-76029-Revision-0

Power System Waveform Datasets for Machine Learning

Christopher Roger Sticht, Stephen Arthur Bukowski

January 2024

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

Power System Waveform Datasets for Machine Learning

By:

Christopher Sticht

&

Steve Bukowski

Introduction

The desire for increased visibility across the electricity grid will necessarily increase the deployment of sensing and measurement devices and associated data management needs to unprecedented levels. For the existing sensing and measurement infrastructure, there remains a great amount of "value" yet to be extracted through advanced data management and analytics.

Availability of more data will not, by itself, lead to changes in grid visibility, security, and resiliency. To create the predictive and prescriptive environment required to enable new markets and transactions for customer revenue and a reliable grid, the data must be collected, organized, evaluated, and analyzed using sophisticated algorithms to provide actionable information allowing operators and customers to reliably manage an increasingly complex grid.

Progress in artificial intelligence (AI) has been largely driven by large, publicly available datasets that can be used to train AI algorithms such as MNIST, a database of handwritten images of digits, and ImageNet, an image database of everyday objects. These types of publicly available databases of real-world training datasets have been largely credited for advancement of image processing, computer vision, and deep learning algorithms that these use cases deploy. However, in the power systems industry to date, there are few databases with proper event labeling, and data access to a publicly available collection of power system event waveforms that will allow users to interact with grid signature data. Publicly available datasets of power system event waveforms, such as the DOE/EPRI dataset, often lack critical metadata or contain limited examples of each event type, and data formats vary widely across these datasets.

Background

Today, the electricity sector faces several challenges in the pursuit of reduction of carbon and shift to renewables. One significant contributor to the challenges is the rapid shift of generation resource characteristics away from the traditional "big iron" generation based upon synchronous rotating machines that are dispatchable and provided large fault currents, reactive power, and inertia for the system to inverter-based resources (IBR) power by renewable energy sources, which as of today have been driven to different design characteristics based upon adaption to the grid. Synchronous generators have provided foundational characteristics of how the grid, or our Electrical Energy Delivery System (EEDS) is operated including areas of reliability, resiliency, system dispatch, system stability, and system protection. IBRs do not share the same physics of the synchronous generator, hence as we continue to deploy more and more IBRs we can begin to erode the operational, system stability, and system protection methodologies used in the electrical energy delivery system over the last 120 years.

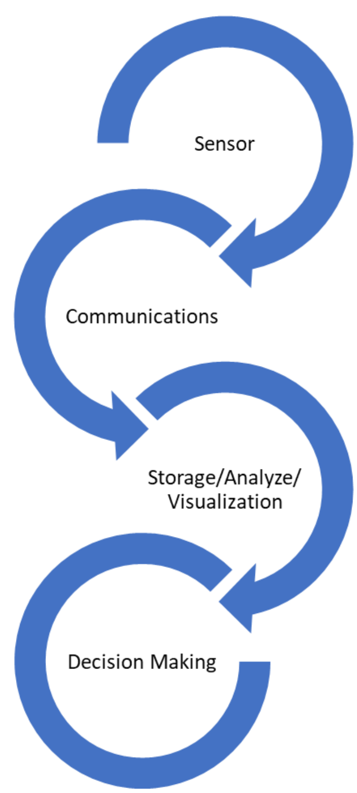

Figure 1 Sensors to Decisions

An example of the change to existing methodologies can be found in system protection with fault current and sequence current assumptions used when identifying a fault. The fault currents from

IBRs are significantly less than traditional synchronous generation sources. To achieve carbon and renewable goals being set today, we must address the challenges of IBR sources now to ensure a clear pathway for a carbon free grid while continuing the reliability, resiliency, and affordable electrical energy in the US.

The evolution of sensors and the associated technology and associated ecosystem components of the business process and decision-making process have had slow adoption within the electric utility industry and continue to bring significant challenges for the electric utility industry as advancements in different precursor technologies are made at different rates, scales, and times, challenging integration, costs, ROI, and functionality. However, these areas of technology are at an inflection point of advancement, providing significant opportunity to evolve the operation of power systems via low-latency, high-bandwidth communications, lower cost of embedded systems, and the growth of large data, machine learning, and artificial intelligence based upon real-time data.

New sources of data and the increased number of recording devices is increasing the amount of data. The convergence of modernization efforts with advances in computing, statistics, machine learning, deep learning, and artificial intelligence is providing opportunities to achieve more efficient, autonomous operation, control, and protection for the grid providing novel core analytic, assessment, and engineering capabilities, while also providing additional depth of defense for cybersecurity postures of electric utilities.

Recently, statistical and machine learning techniques (MLTs) have proven to be effective in numerous applications including power system studies. Various MLTs such as artificial neural networks (ANN), Decision Tree (DT), support vector machines (SVM) have been proposed, resulting in effective decision making and control actions in the secured and stable operations of the power system [1]. Further work advancing MLT in power systems has been done transient stability, protection, and control (Transient Stability Assessment-TSA) [2] [3], have provided work into demonstrating MLTs in the fault identification and areas of protection, which justify extending this work into real-time applications and improved upon for constant monitoring.

It should be noted that the recent Bipartisan Infrastructure Law (BIL) Solar and Wind Grid Services and Reliability Demonstration is seeking solutions related to this research and target outcomes and recognizes the emerging challenges to protection and power systems operation.

Figure 1 shows the curve for hits on the google scholar search term “Machine Learning in Power Systems,” which clearly demonstrates the exponential increased popularity of MLTs into power systems.

Machine learning in power systems today is based upon static files which is significantly different than data types needed for “real-time” ML applications. Additionally available data is limited and sparsely available.

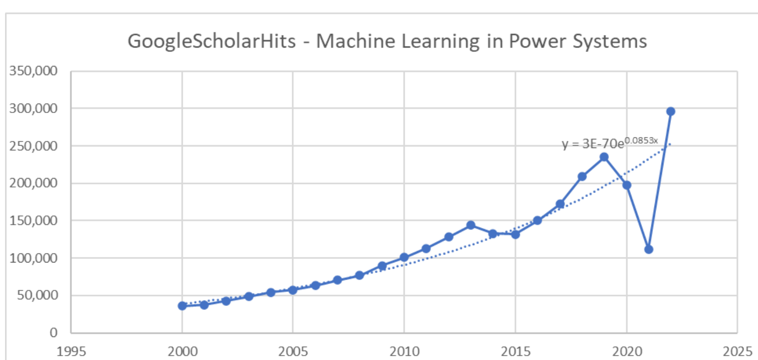

Figure 2 Google Scholar Hits May 2022 for Machine Leaning in Power Systems

All MLTs require training data and the quality of the data is critical to its performance, accuracy, and complexity of the algorithm. Power grid data sources can vary significantly based on the objective of the machine learning (ML) algorithm. Data for Power Systems Operation (PSO) centers around the physical electrical properties associated with power flow and energy delivery. Today a significant portion of electric utility's primary data for operation decisions come from a few types of devices (protective relays, RTU's, revenue meters) over different time scales.

Measuring and recording devices are connected throughout the electric power system and represent different basis for processes within the utility and include but not limited to Voltage (V), Current (I), real-power (P), reactive-power (Q), frequency (f), phase angle (Φ), time (t) and derivatives of P and Q over time including Energy (kWh/MWh).

Data sets for machine learning in PSO research generally originate from simulated data from real-time simulation software or data sets available from historical capture of utility SCADA data, meter data, relay fault records, or synchrophasor data. This data is often referred to as off-line data and packaged into training data sets for different MLTs. Each of these sources of data have challenges for MLTs. Challenges include limited availability and tend to be unstructured, multi-modal, heterogeneous, decentralized, and highly nonlinear in nature based upon the owner of the data and decisions around timing, selection of data type, and locational limits.

Data sets geared towards transients and stability are challenged by high variety, volume, velocity, and low veracity, [4] [5] describe these characteristics in the following manner:

1) Data variety – multiple sources of unstructured and semi-structured data that typically require preprocessing required to derive useful meaning and supported metadata
2) Data volume – high volume, number of devices x sample rate x duration
3) Data velocity – rate of sample for the data
4) Data veracity – ratio of meaningful data to non-meaning full data represented for example by the seldom occurring faults versus duration of normal operations

In addition to these challenges for ML data sets, MLTs that focus on transient stability problems, protection, and control can require granularity supported by higher sampling rates for which available datasets are severely limited.

As identified by [4], a key reservation for many power systems sensors is that the data is either not logged or is overwritten very quickly as in protection relays and related sensors. Additionally, relays must be pre-programmed to detect and record and event which has a few cycles of data preceding the event. Other available types of data for ML work in transient stability, protection, and control problems must be based upon RTU/Historian data that is 2-6 second interval and synchrophasor data at a maximum rate is one sample per cycle, which can be insufficient for the time period of the associated work and often filters and may lose information based upon RMS and Discrete Fourier Transforms (DFT). Essentially high-fidelity data from multiple locations in the power grid to support MLT in transient stability, protection, and control is extremely limited.

Datasets for Machine Learning

The Google Speech Library is a good example of how large datasets have helped advance machine learning technology. The Google Speech Library has approximately 2000 samples of each spoken word. This is the beginning of how Google can provide speech recognition services. Here is a sample of spoken word waveforms, can you identify which word is represented by the unlabeled waveform by comparing it to the three single samples of the waveforms for the words bed, cat and off?

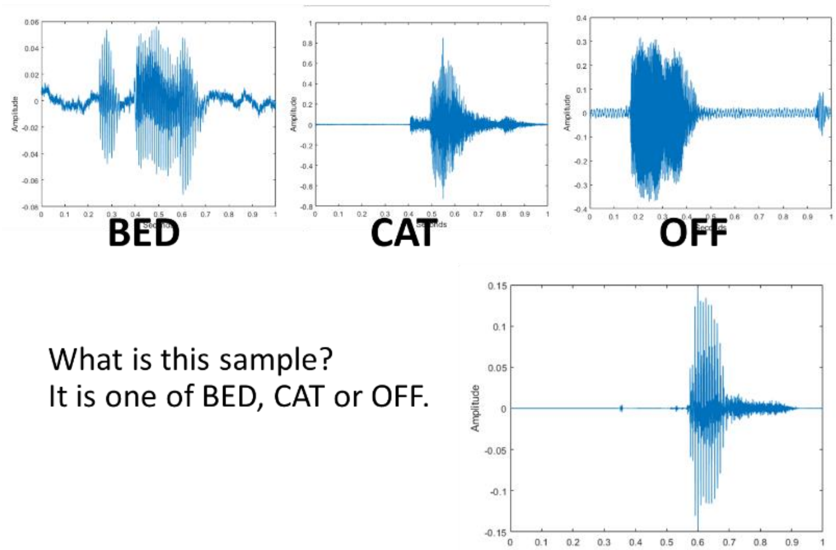

Figure 3 Example of Audio Data for Machine Learning

Below are three samples of the unknown word spoken by different people. Can you see the similar pattern for the unlabeled sample is from the expanded example set? This is why so many samples of each word is needed.

Machine learning requires variety in example data to learn the patterns of what it is being trained to look for. In this example it was spoken words. The Google example has been repeated for things like handwriting recognition, heart cardiogram recognition, various forms of motion sensors, image recognition, etc.

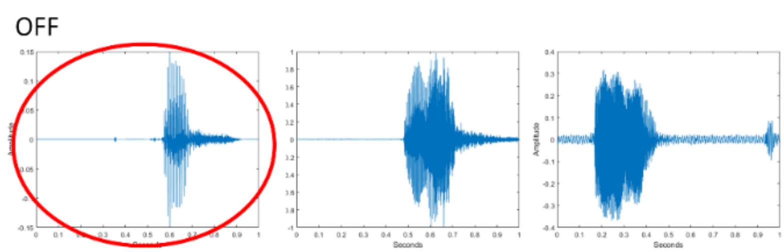

Figure 4 Audio Waveforms for Word OFF

Waveform Data

The power system delivers Alternating Current (AC) power to electricity consumers. The AC power is a sinusoidal signal and is measured at many locations across the power grid. When different events take place on the power grid unique waveform signatures are created during the different events as illustrated in the image below.

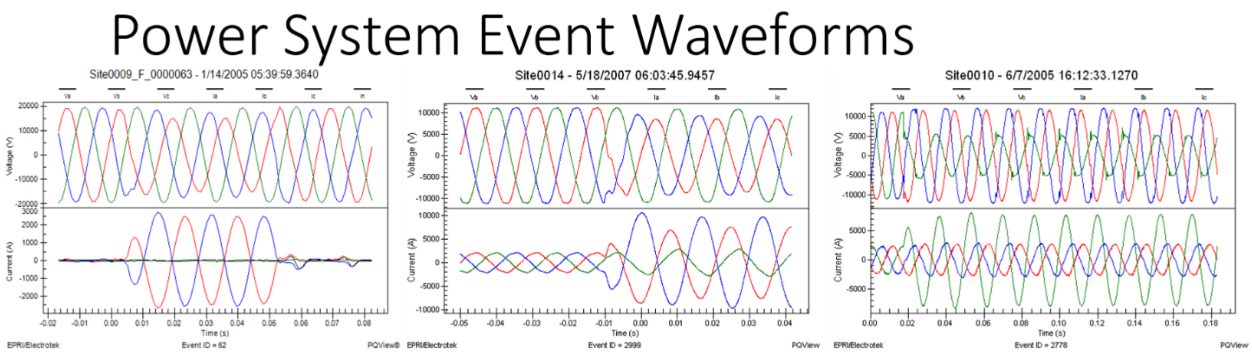

Figure 5 Examples of Power System Event Waveforms [18]

Researchers do not yet have an organized and labeled power system waveform data set to use for neural network training. The most comprehensive data set for power systems analysis is the PMU data set for FOA 1861 (https://www.naspi.org/). The PMU data is at a sample rate of 30 to 60 samples per second which is insufficient for detailed waveform representation particularly harmonic characteristics. Using waveforms as a direct source to analyze system events requires higher resolution than PMU data provides. High-resolution event recordings are highlighted by [6] and has been implemented in some power systems [7]. Point-on-wave (PoW) data, features a sample rate of 1 kHz or higher.

The DOE/EPRI National Database Repository of Power System Events (https://pqmon.epri.com) contains approximately 300 event recordings with event type classifications/labels such as tree, weather, equipment, etc. Each event recording includes event files that include images of waveforms and raw datafiles in various formats. Each event also includes event properties with more detail

Recent work to improve data availability include the DOE Grid Event Signature Library (GESL), which aims to develop a collection of labeled events or anomalies that is developed to enable statistical and machine learning and traditional analytics research to predict and monitor power grid health. The GSL currently has datasets from ten different sources with multiple events/signature and sampling rates with intentions to continue to expand. The GESL does a good job of developing this open-source database framework but is limited by the availability, accurate labeling, real-time quality, and source of data, specifically in ML for transient stability, protection, and control.

Existing grid event datasets are inadequate for AI/ML applications

- Not enough samples of each event type to train into AI/ML algorithms
- Non-existent or unlabeled events
- Non-standardized sampling rates limits integration of datasets
- Inconsistent sample rates for different fields within the same event recording
- Low sampling rates cannot detect transient events
- Precursor events are not tracked and stored

Power Systems, Machine Learning & Neural Networks

Several efforts have started to examine machine learning capability in the areas of transient stability, protection, and control, specifically the functionality of fault identification, [2] [3] through MLT, but have not yet addressed the real-time monitoring aspect and the performance envelope of relaying.

Some work in the industry has investigated phase identification [8] [9] [10]. There was not a lot found in the literature where voltage and current waveforms have been used. This may be partially due to the relative lack of readily available waveform data for research.

Some techniques have been developed that use contextual information [11] [12] such as weather, affected phase(s), season, event time, and interrupting device. Other techniques use extracted features from waveform data [13] [14] such as the derivative of current and voltage signal, energy, amplitude, correlation coefficient, etc. Self-recoverability, zero current time, degree of distortion, transition time and waveform randomness, are extracted from the recorded waveforms in [15], and used as the input to a fuzzy inference.

AI/ML for Protection

The value of waveform data used in Neural Networks allows for the advancement of protection for the power grid. Ultimately, plug and play protection and self-healing on the power system is the goal.

Today, protection engineers choose protection settings that define protection curves on a Time-Current Curve (TCC) (see Figure 6). This approach works for radial circuits with strong utility sources. Long circuits may exhibit low available fault current at the end of the line.2 Additionally, modern distribution systems with Distributed Energy Resources (DER) such as wind, solar and batteries are becoming more prevalent. In these and other contexts, the conventional approaches to protection are facing new challenges.

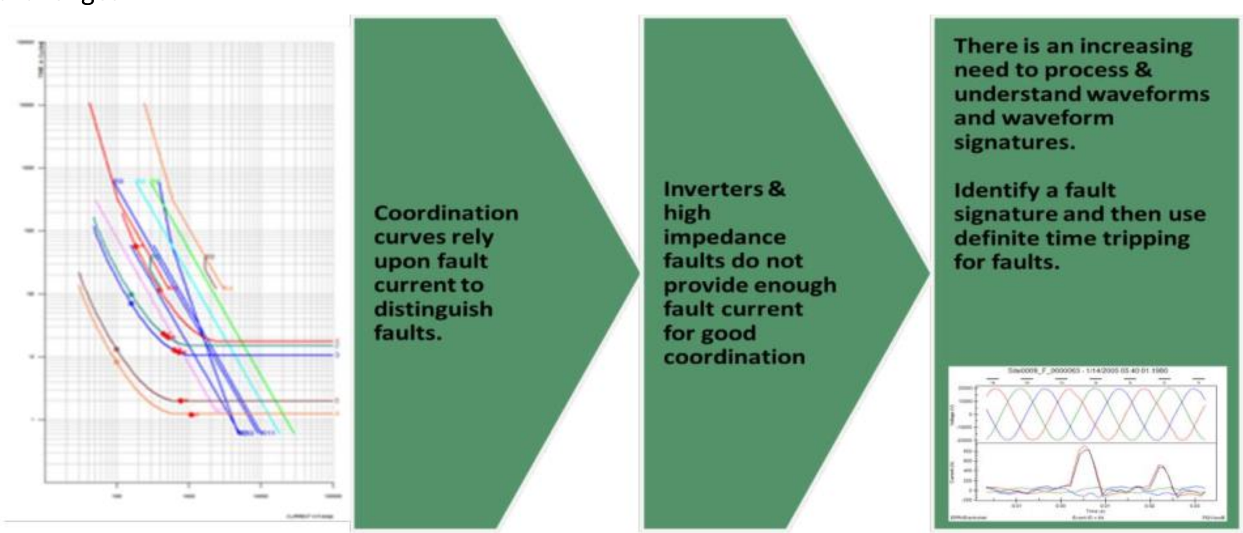

Figure 6 Evolution of Protection with Waveforms and Waveform Signatures

Power System Waveform Processing

Many modern devices and systems record waveforms of power system events. In some cases, utilities may record a half-a-dozen event captures for every event. This is thousands of events per year. The industry will need faster, more automated, more conclusive, and easier to use systems going forward that can process massive amounts of event recordings without extensive input/support from power system engineers.

The fields of audio engineering, speech and audio processing use spectrograms fed into MLTs to perform tasks such as word recognition, speaker recognition, and cleaning noise from the signal of interest. These are all examples that have direct equivalents in power system waveforms and waveform "signatures".

Audio signals are very similar to power system signals other than the fact that the frequency of audio signals tends to be high (100 to 20,000 HZ) where is the frequency of power system waveforms tend to be low (~60Hz). Additionally, audio waveforms are often singular multi-frequency signals which result in a single waveform, whereas power system waveform data exists in three phases of current data and three phases of voltage data, yielding six waveforms altogether.

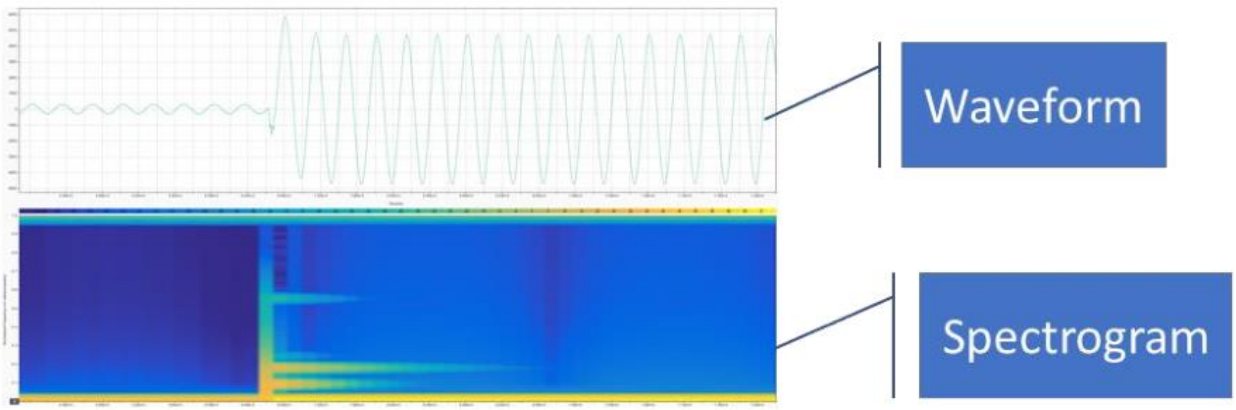

Figure 7 Spectrogram of a Waveform

In audio signal processing, the audio waveform data is converted to an image using spectrograms [16] [17]. Spectrograms are graphical representations of signal in the frequency domain, or Fourier transforms. Short time Fourier transforms are time sampled Fourier transforms of waveform data. In other words, a Fourier transform is performed on a snippet of waveform data (a time step). The next snippet of waveform data is also converted to a Fourier transform and so forth and so on until the entire waveform recording is converted to samples of Fourier transforms. Each Fourier transform results in a harmonic spectrum of intensities for each harmonic frequency that exists within that snippet.

Each of the snippets has a harmonic spectrum associated with it that describes all the frequency components of that snippet. The frequency spectrum of each snippet is horizontally stacked next to each other for all snippets of the overall waveform. This creates a two-dimensional array of data where each column represents a snippet of time within the waveform and each row represents a harmonic component (see Figure 7).

The array of values can be interpreted as an image where at or near the bottom of the image are the low frequency values and the of the top the image are the high-frequency values of the harmonic components for each time step (see Figure 9).

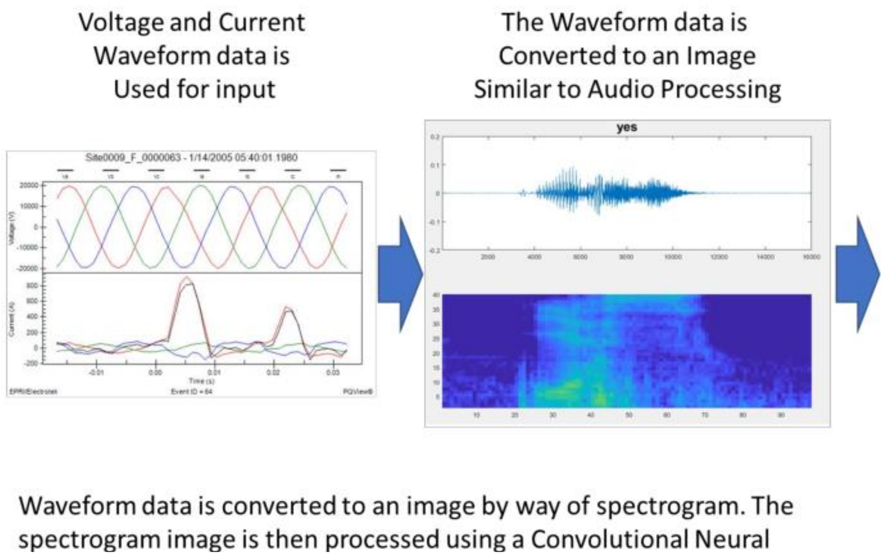

Figure 8 Convert Power System Waveform to a Spectrogram Image Similar to Audio Processing

This image of the Short Time Fourier transform can be passed into a MLT. In the case of power system data some additional modification of the data is required before passing it to a MLT for training (see Figure 10).

A waveform signature could be treated like a spoken word such as “YES” or “UP”. No two people produce the exact same waveform when speaking each of these words, but the algorithm can still distinguish the word regardless of the speaker. No two circuits will produce the exact same waveform for a given fault type (phase-to-ground for example), but the MLT can be trained to classify the fault type regardless of the circuit or location on the circuit.

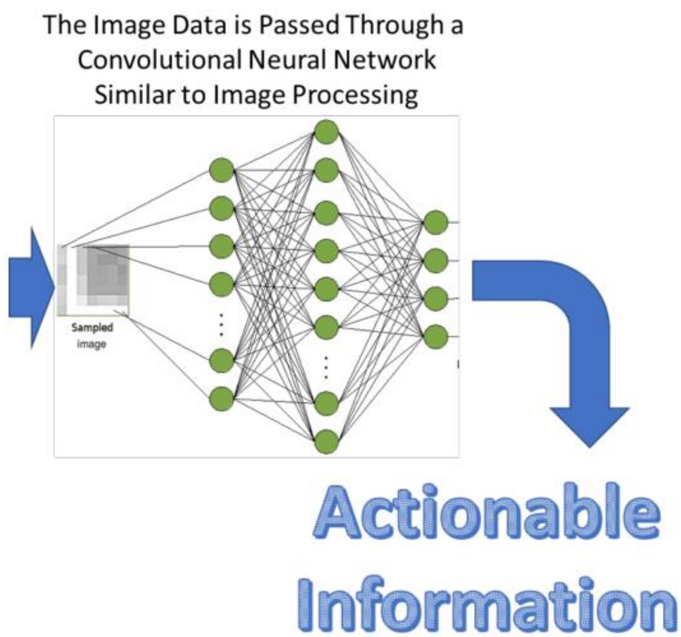

Figure 9 Pass the Spectrogram Image Through a Convolutional Neural Network to Extract Actionable Information

Power system waveform data such as that in Figure 9 can be processed using MLTs. The voltage and current waveform data are used for input into an MLT. The MLT reduces the input data to actionable information (see Figure 10).

Power System Waveform Data for Neural Network Training

To fully represent the waveform, including intermediate and higher order harmonics, high sample rates are required. The waveform data captured by power quality meters and other modern instrumentation typically has sample rates of 3000 to 20,000 samples per second. The Power System Neural Network (PSNN) [3] that leverages spectrograms and CNN is designed for these high sample rates.

Waveform data is becoming increasingly available throughout the industry from power quality meters, microprocessor-based relays, event recorders and other intelligent electronic devices (IED). However, at this point researchers do not yet have an organized and labeled power system waveform data set to use for neural network training.

Given that large and comprehensive publicly available waveform data sets are not yet available to researchers, other approaches will need to fill the gap until the data is available.

While the event waveform captures are being recorded for a wide variety of events throughout the industry, at this point many of the events are unlabeled. In other cases, the event waveform captures do not exist yet. Until the industry gets to a point where event waveform captures from the field can be used for neural network training simulated data will need to fill the gap in the interim.

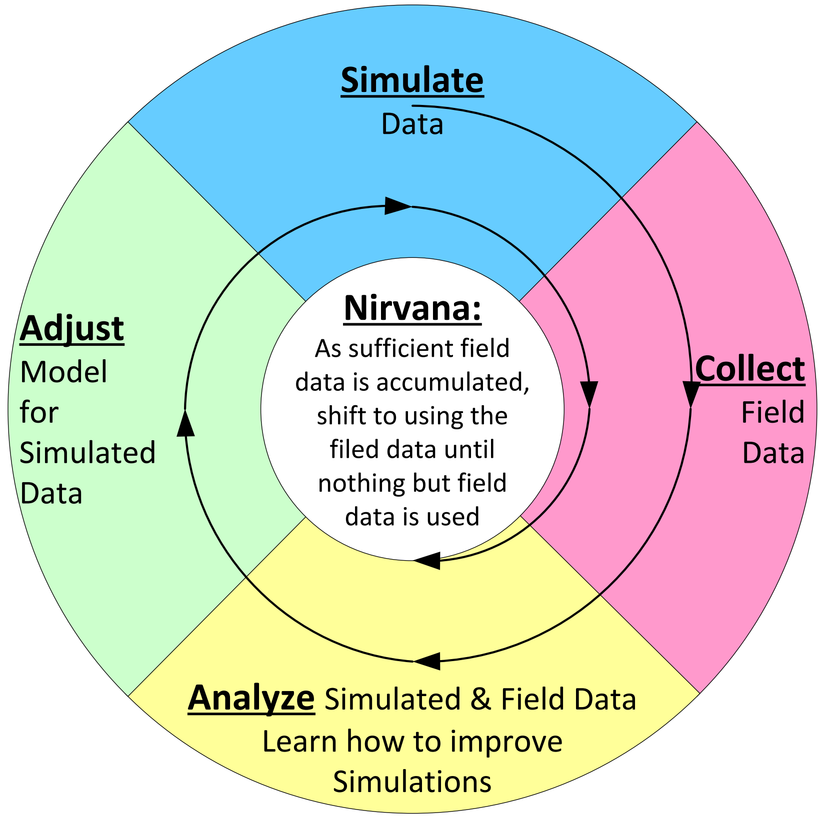

Figure 10 Data Cycle

Simulated data often consists of idealized examples of events and may not include characteristics from field conditions. Therefore, simulated data can be used as a starting point, however simulation tools will need to be improved as time progresses to generate waveforms that more closely reflect field conditions. To that end, we propose here a process by which waveforms are simulated and refined. As the simulated data is being used field examples are collected. As these field samples are collected, they are compared against the simulated waveforms. With understanding gained from this comparative analysis, the simulation models are adjusted to create more realistic simulated waveforms. This cycle repeats, allowing for new simulated data to be created with the improved simulation tools. The cycle repeats until a point is reached at which there is sufficient field data to successfully train the neural networks and simulated data can be shelved or used for more rudimentary training purposes (see Figure 11).

Basic, Time-synched and Geographically Dispersed Data Sets

When it comes to waveform event capture data, there is very little data available to researchers to develop machine learning algorithms. A classification and labeling scheme has been developed for waveform event captures as part of the GSL project. Most event types are simple event datasets for machine learning of single distribution feeder types of events that do not require recordings from multiple locations across the system and to have those events time-synched with each other or for the events to be labeled for the location on the power grid or any other identifying characteristics.

This is significant, as many utilities are reluctant to share event data if it can be identified in any way back to their system. To get field data from utilities it will be critical to develop a methodology to sanitize datasets so that the data cannot be traced back either in whole or in part to the utility who provided the data. This will require working with utilities, research engineers and legal experts to determine how to sanitize the shared data and to create a legal framework for anonymity in sharing data.

The next level of data that will be needed is data that is tied to physical systems. This might be referred to as a time-geo dataset for machine learning. For example, waveform data from multiple sensors at multiple locations on a transmission system to help identify what a particular event looks like from different locations on the same power system. Some investigations have already started in this direction such as investigating capacitor switching evens from multiple locations on a power system to be able to conclusively determine the location of the event from various remote locations [18]. Field data to support this type of study requires information of the actual system such as what might be found in a power flow model. Further, the event recordings would need to be time-synched with each other and the locations of where the readings were taken need to be labeled.

Dataset types for Power System Machine Learning Algorithms

1. <u>simple event datasets</u>
2. <u>time-geo dataset</u>
3. <u>wholistic power system dataset</u>

This kind of dataset is far more complex and involved. The sanitization of this type of dataset is similarly more complex and would require a more in-depth investigation into the legal and liability aspects of sanitizing and sharing such a dataset with researchers in a broad sense.

The most complex form of data set would be like the time-geo dataset for machine learning, but also requires data concerning control system information, communication system data and cyber-security data. This might be referred to as wholistic power system dataset for machine learning. This type of dataset would be required for cyber-security machine learning algorithms, power system operations machine learning algorithms and power system communication system machine learning algorithms.

Data Labeling

A critical component for the use of power system waveform event recording data files is for each file to be accurately labeled with respect to the event that is taking place in the recording. Labeling may take different forms including:

- Label(s) for the entire file
    - There can be multiple labels for any given recording such as the type of fault and the phase that the fault took place on. (lightning stroke on phase A)
- Label(s) for each timestep in the file

Supervised learning neural network algorithms depend on accurate labels for the data. The type of label (file-level or timestep-level) depends on the algorithm of interest. Some machine learning structures/approaches require file-level labeling while others require timestep-level labels.

Labeling is not automatically included in field data. There are ways to expedite the labeling of event recordings. One approach to expedite some labeling at a file level is to cross-reference the timestamp and equipment identification of each event recording file against the time-step and equipment identification of the Outage Management System (OMS) records which includes the event label. However, there are many events that will have event recordings that are not included in OMS records such as standard switching events.

At this point, timestep-level labeling, and many other events (such as switching) requires manual labeling in much the same way that image labeling is done for video and image labeling is done. In this case, each event recording file must be manually checked by an engineer or similarly trained technician to correctly identify the event taking place in the event recording. There are some algorithms developed (usually in-house algorithms at utilities) to expedite the process, but to date there is no commercially available tool to automatically label power system event recording files. In fact, that highlights the need for labeled power system event recordings. The labeled event recordings will be used to help develop the machine learning algorithms and tools that will later be used by the industry to identify the type of event seen in new event recordings.

Where is the Industry Going?

Datasets are an integral part of the field of machine learning. Machine learning algorithms function by training on a large set of example data. High-quality labeled training datasets for supervised and semi-supervised machine learning algorithms are usually difficult and expensive to produce because of the large amount of time needed to label the data.

The future of power systems will likely follow a similar path to the other applications of Machine Learning and Artificial Intelligence. It needs to start with a similar library of waveform event captures for a wide variety of power system events. Initially the event recordings will largely be hand-annotated and will need to contain 50,000 to 100,000 event recordings. Where each event type will consist of several hundred event recording files. All of which needs to be normalized, organized, and categorized.

The dataset will be applied to machine learning research in power systems. For many academic and industry researchers, the availability of truth-labeled test data helps drive algorithm research. The power systems dataset and labeling problem is a much more challenging problem then the audio industry for voice recognition or image recognition, as access to the resolution of data needed as well as the labeling is not available and needs to be developed. Interviewing 1000 people for voice recordings or taking the output of digital camera is significantly lighter lift to achieve than power system waveform data for different types of events under transient and steady-state conditions.

Summation

Today, the electricity sector faces many challenges. Renewable generation, independent power producers, and power purchase agreements are changing the landscape in Transmission & Distribution (T&D) system operations, planning, protection, and generation/storage. This is raising questions on some of the fundamental premises that engineers have relied on for several decades.

There is an increasing availability of sensing and measurement devices and associated data across the T&D system. More data will not, by itself, lead to changes in grid visibility, security, and resiliency. The data must be collected, organized, evaluated, and analyzed using sophisticated algorithms to provide actionable information allowing operators and customers to reliably manage an increasingly complex grid.

Machine Learning in Power Systems is becoming a significant area in which solutions for future power systems are being developed. Machine Learning has been largely driven by large, publicly available datasets that can be used to train sophisticated algorithms. In the power systems industry to date, there are insufficient collections of waveform data with proper event labeling that is publicly available for researchers to develop the algorithms needed for the grid of the near future.

There is a powerful need for datasets of waveform event capture data that is sufficiently large, organized and labeled for machine learning researchers to meet the needs of today's power system researchers. The place to start is:

1. Develop data sharing, data securing, data sanitizing processes and agreements for electric utilities to share their event recordings (safely/anonymously) with the research community
2. Develop simulation models and algorithms to generate simulated datasets
3. Acquire field data from utilities and other sources where available
4. Develop techniques to update simulation models and algorithms to more accurately reflect field data

Labeling is a critical component of the need for power system event waveform files. Without accurate labels, the waveform files are of little use. This component is time consuming, expensive, and labor intensive but essential for a useful dataset for research.

Bibliography

[1] O. A. Alimi, K. Ouahada and A. M. Abu-Mahfouz, "A Review of Machine Learning Approaches to Power System Security and Stability," *IEEE Access,* vol. 8, pp. 113512-113531, 2020.

[2] A. Thomas, S. Koshy and R. Sunitha, "Machine Learning Based Detection and Classification of Power System Events," in *2020 International Conference on Power, Instrumentation, Control and Computing (PICC)*, 2020.

[3] C. Sticht, "Power System Waveform Classification Using Time-Frequency and CNN," Oak Ridge National Laboratory, Oak Ridge, TN, 2022.

[4] H. Akhavan-Hejazi and H. Mohsenian-Rad, "Power systems big data analytics: An assessment of paradigm shift barriers and prospects," Energy Reports, 2018.

[5] H. Yang, D. Shang, T. Chen and C. Li, "Machine learning for power system protection and control," *The Electricity Journal,* vol. 34, no. 1, 2021.

[6] A. a. J. F. Silverstein, "High-resolution, time-synchronized grid monitoring devices," NASPI (North American Synchrophasor Initiative), 2020.

[7] EPB Chattanooga, *System Wide S&C IntelliRupter® Waveform Captures, Shared under NDA.,* 2020.

[8] M.-F. N.-C. Y. a. W.-F. C. Guo, "Deep-Learning-Based Fault Classification Using Hilbert-Huang Transform and Convolutional Neural Network in Power Distribution Systems," *IEEE Sensors Journal,* vol. 19, no. 16, p. 6905–6913, 2019.

[9] K. a. P. W. P. Zhu, "Fault Classification of Power Distribution Cables by Detecting Decaying DC Components with Magnetic Sensing," *IEEE Transactions on Instrumentation and Measurement,* vol. 69, no. 5, p. 2016–2027, 2019.

[10] V. a. A. Y. Ashok, "Fault Diagnosis Scheme for Cross-Country Faults in Dual-Circuit Line with Emphasis on High-Impedance Fault Syndrome," *IEEE Systems Journal ,* 2020.

[11] X. B. S. a. S. M. Jiang, "Automated Distribution Network Fault Cause Identification with Advanced Similarity Metrics," *Transactions on Power Delivery,* vol. 36, no. 2, pp. 785-793, 2020.

[12] M. S. A. S. a. M. R. Bashkari, "Outage cause detection in power distribution systems based on data mining," *IEEE Transactions on Industrial Informatics,* vol. 17, no. 1, pp. 640-649, 2020.

[13] M. A. H. S. a. T. G. Jarrahi, "Novel change detection and fault classification scheme for ac microgrids," *IEEE Systems Journal,* vol. 14, no. 3, pp. 3987-3998, 2020.

[14] A. E. F. E.-S. a. M. S.-P. Hooshyar, "Fault type classification in microgrids including photovoltaic DGs," *IEEE Transactions on Smart Grid,* vol. 7, no. 5, pp. 2218-2229, 2015.

[15] Y.-L. e. a. Liang, "Typical fault cause recognition of single-phase-to-ground fault for overhead lines in nonsolidly earthed distribution networks," *IEEE Transactions on Industry Applications,* vol. 56, no. 6, pp. 6298-6306, 2020.

[16] V. Velardo, "Audio Signal Processing for Machine Learning," YouTube, 2020.

[17] M. Müller, Fundamentals of Music Processing, Springer International Publishing, 2015.

[18] A. Wilson, H. Tran and D. Lu, "Uncertainty Quantification of Capacitor Switching Transient Location using Machine Learning," *IEEE Transactions on Power Systems,* 2023.

[19] DOE/EPRI, *DOE/EPRI National Database Repository of Power System Events.*