Clustering and Dimensionality-reduction Techniques Applied on Power Quality Measurement Data

Gjert H. Rosenlund
Dept. of Energy Systems
SINTEF Energy Research
Trondheim, Norway
gjert.rosenlund@sintef.no

Kristian W. Høiem
Faculty of Science and Tech.
NMBU
Ås, Norway
kristian.wang.hoiem@nmbu.no

Bendik N. Torsæter
Dept. of Energy Systems
SINTEF Energy Research
Trondheim, Norway
bendik.torsater@sintef.no

Christian A. Andresen
Dept. of Energy Systems
SINTEF Energy Research
Trondheim, Norway
christian.andresen@sintef.no

***Abstract*—The power system is changing rapidly, and new tools for predicting unwanted events are needed to keep a high level of security of supply. Large volumes of data from the Norwegian power grid have been collected over several years, and unwanted events as interruptions, earth faults, voltage dips and rapid voltage changes have been logged. This paper demonstrates the application of clustering and dimensionality-reduction techniques for the purpose of predicting unwanted events. Several techniques have been applied to reduce the dimensionality of the datasets and to cluster events based on analytical features, to separate events containing faults from a normal situation. The paper shows that the developed predictive model has some predictive capability when using balanced datasets containing similar muber of fault events and non-fault events. One of the main findings, however, is that this predictive capability is significantly reduced when using unbalanced datasets. Thus, the development of an accurate predictive model based on normal power system conditions, i.e. an unbalanced dataset of events and non-events, is a topic for further research.**

***Index Terms*—Machine Learning, Unsupervised Learning, Power system, Power Quality Analysis, Fault Prediction, Dimensionality-Reduction**

I. Introduction

A. Motivation and Background

The introduction of ever-increasing amounts of intermittent renewable generation, coupled with the increasing electrification of the European societies, leads to an increased strain on the power grid and its operation. In order to maintain high security of supply, it is paramount to evolve the tools used for power systems operations. One such tool would be the ability to predict undesired events with sufficient prediction horizon to facilitate mitigating actions. Development of such a tool is spurred by recent advancements in data-driven techniques, machine learning (ML), available data volumes and computational resources.

B. Relevant Literature

Even though machine learning techniques are widely used in other areas of power grid management, active work on predicting fault events using power quality (PQ) data is quite sparse. Interesting work on the topic include [1]–[4]. [1] utilize μ-phaser measurement units (μ-PMU) data and semi-supervised learning for event detection and classification, [2] cluster momentary faults that are recurring in the grid, for example vegetation touching the overhead power line. [3] combine an artificial bee colony algorithm with neural networks to find optimal features for classifying disturbances in a simulated data-set and [4] use anomaly detection techniques to identify anomalies in PQ time-series.

C. Contributions and Organization

This paper presents work done within predictive methods utilizing clustering and dimensionality-reduction techniques on large datasets of PQ measurements from the Norwegian power system. A number of techniques has been applied, and the algorithmic performance on balanced and unbalanced datasets have been investigated. Using two datasets ensure a balanced set for model development, and a realistic one-node sequential dataset for validation of how the algorithms work in an operational setting.

This paper address the knowledge gap in the joint domains of power systems and data science. More specifically, this paper pursue the state-of-the-art in using power quality data and unsupervised learning algorithms to group fault and non-fault events based on analytical derived features.

Chapter II describes the input data. In Chapter III, the methodology used in the analyses is presented. In the following two chapters, Chapter IV and V, the results from the investigations are presented and discussed. Finally, the conclusion and suggested further works are presented in Chapter VI. It is the desire of the authors that other groups working in this young but active field will learn from the results of this study, and be inspired to surpass our results.

II. Data

The authors have been granted conditional access to power quality data for the majority of the Norwegian power system by a number of distribution system operators (DSOs) and the Norwegian transmission system operator (TSO). The underlying database utilized in this paper spans the period from January 2009 to early March 2020. The nominal line voltages varies from 10 to 420 kV. A total of roughly 270

The authors would like to thank the Research Council of Norway and industry partners for the support in writing this paper under project 268193/E20 EarlyWarn.

This is the accepted version of a paper published in 2020 International Conference on Smart Energy Systems and Technologies - SEST
http://dx.doi.org/10.1109/SEST48500.2020.9203294

years of PQA data has been collected from 49 measurement nodes, giving on average 5-6 years of historical data from each node. However, the number of years of available data varies significantly from node to node. The data owners regard sharing of data as a joint effort to strengthen the research within the field, a mutual beneficial endeavor.

1) Data sources: This paper exploits data from power quality analysers (PQAs). The paper analyses data from Elspec PQAs, which continuously sample voltage and current waveform at a sampling frequency of up to 50 kHz, with data being compressed where appropriate. The operational Elspec devices collect and compress many events and disturbances each year, and some nodes have been online for over 10 years. To properly manage and extract value from such a massive dataset, two software packages have been developed. The Automatic Event Analysis (AHA) program is used to automatically detect and report lists of fault events and disturbances in the time series from the PQAs [5]. The tool can identify and classify interruptions, earth faults, voltage dips and rapid voltage changes. These are annotated with event type, start time and end time for each event.

A large portion of applications and algorithms within machine learning requires labeled datasets for exploitation of patterns and signals in the data, called *supervised learning*. These labels are extracted using AHA, mentioned above. To detect explanatory signals for predictive purposes, one also need the power quality data leading up to the event. For this purpose, the Dynamic Dataset Generator (DDG) software has been developed. For details on DDG, see [6], [7]. This program takes a fault event as an input and extract user-specified variables, such as voltage and current waveforms, harmonics and RMS values at the resolution and duration requested by the user.

The analysis presented in this paper revolves around two datasets. The first dataset consists of in total 4579 events, whereas 2294 of these are fault events. Fault events include voltage dips, earth faults and interruptions. This dataset is termed the *balanced dataset*, since the number of faults to non-faults are approximately balanced. The second dataset consists of sequential data from one node in the Norwegian power grid. This dataset consists of 66240 samples, where 22 are fault events. The second dataset is therefore labeled the *unbalanced dataset*. Utilizing the relevant machine learning models requires a large volume of data for each class. It is the purpose of the balanced dataset to serve as a dataset for experimentation, while the operational performance of any models is tested by running the models on the unbalanced dataset, simulating a real-time performance of the algorithms.

2) Data pre-processing: When presented with the data, the most appealing approach is to leverage algorithms that are tailored for time series forecasting, and use the raw data as explanatory variables. This approach is investigated in [5]–[8].

In the proposed approach, manual feature engineering is conducted, leveraging expert human insights and reasoning. The engineering of features is based on discussions with industry experts and research scientist, bridging the gap between industry and academia. The hypothesis is that when there is a very short, rapid change in the harmonic components in the signal, there is a larger than normal probability for a fault occurring. To represent this signal, no averaging is done. The maximum and minimum values of the line and phase voltages are extracted, as well as the total harmonic distortion (THD) and the 1st-25th harmonics of the voltages. These extracted for each second in the hour leading up to the fault, resulting in a dataset of 3600 by 300 matrix of explanatory variables for each event. To further refine these data, only the maximum, minimum, standard deviation, signal-to-noise ratio and the number of outliers for the time-series are kept, compressing the dataset to a 1500 by 1 vector of explanatory variables. The large number of explanatory variables is the main motivation for exploiting techniques for dimensionality reduction. The data mining process is illustrated in Fig. 2.

III. Method

In previous work [5]–[7], the authors investigate other methods for predicting faults, including supervised learning. They all conclude that it is possible to predict an event of either voltage dip, earth fault, rapid voltage changes, or power interruptions, with a given prediction horizon. To complement these results, unsupervised learning has been investigated. One of the major benefits of using unsupervised learning is the nature of the algorithm to find patterns in the data without the need of labels.

A. Dimensionality-reduction

In the area of analysing multivariate datasets, a common practice is to apply dimensionality-reduction techniques to reduce the dimensions of the dataset, to remove correlated features and to reduce noise [9]. Maaten et. al [10] conducted a comparative review on available dimensionality-reduction techniques. It concluded that even though linear techniques cannot adequately handle complex non-linear data, non-linear techniques for dimensionality reduction are not capable of outperforming traditional linear techniques, such as principal component analysis (PCA). The main function of PCA is to decompose the data into components of explanatory variance, where each sample can be represented in relation to these components.

Another technique widely used in visualising high-dimensional data is t-SNE [11], [12]. This technique minimises the Kullback-Leibler divergence between a high-dimensional point distribution and a proposed low-dimensional point distribution to find a suitable projection of the original points onto the lower dimension. The t-SNE has mainly two hyper-parameters that can be tuned when training; learning rate and perplexity. The learning rate defines the size of the update step in the optimization process. Perplexity is related to the number of nearest neighbours a sample has [11]. To make the figure text more readable, learning rate is denoted as *lr* while

This is the accepted version of a paper published in 2020 International Conference on Smart Energy Systems and Technologies - SEST
http://dx.doi.org/10.1109/SEST48500.2020.9203294

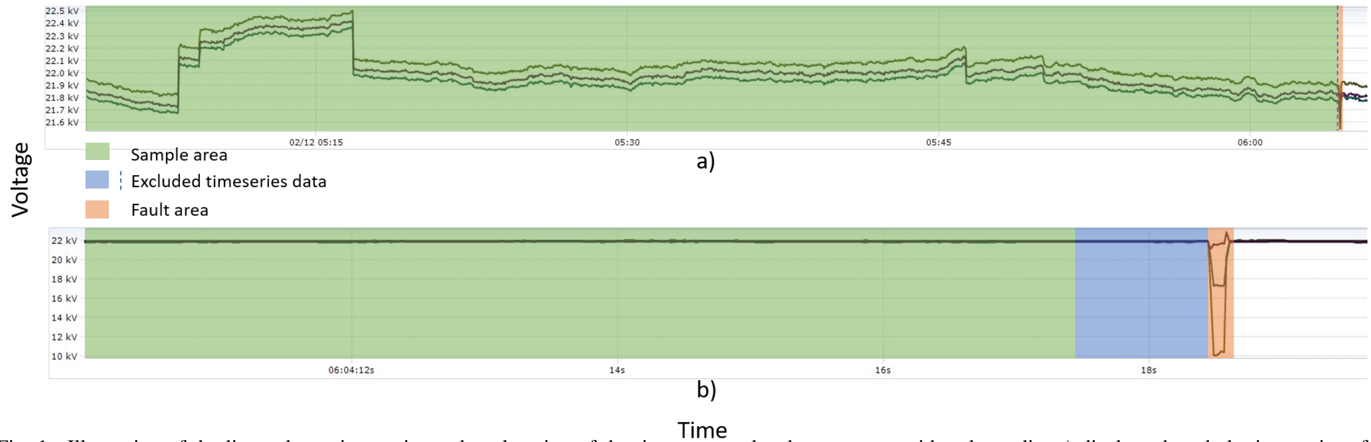

Fig. 1. Illustration of the line voltage time series and explanation of the time range related to an event with voltage dip. a) displays the whole time series of 60 minutes including the voltage dip. b) displays a 10 seconds view of the same sample as in a).

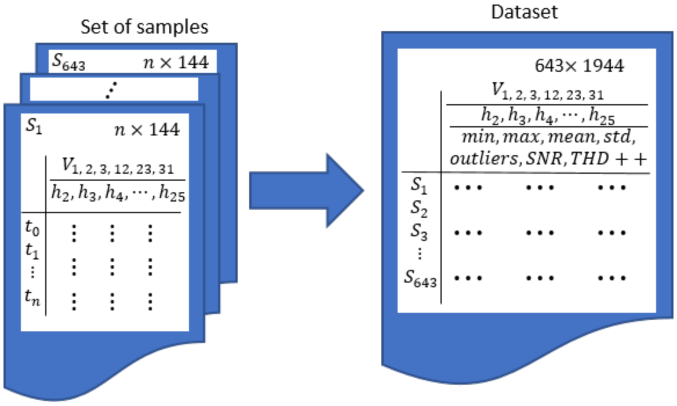

Fig. 2. Transformation of time series sample features of two dimensions into a one-dimensional sample by feature engineering. The transformed samples are to the right represented as a dataset where the columns are a combination of all the elements in the three phases.

perplexity is denoted as p. The number of iterations is denoted as ni.

B. Clustering

Clustering is the division of data into groups of similar objects [13]. It is for the most part an unsupervised technique, where the hidden patterns in the data are explored [14]. Multiple clustering techniques exist today [15]. Jani et. al [16] point out that the k-means clustering algorithm is still widely used despite the fact that it is over 50 years old.

For time series clustering there may be several approaches, such as raw-data-based or feature-based [17]. The feature-based approach is pre-processing the data before conducting the clustering. In the work in this paper, both approaches have been investigated using k-means clustering.

IV. Results

In this section, the results from three different investigations on the same dataset are presented.

A. t-SNE

The balanced and the unbalanced datasets were investigated separately by applying t-SNE dimensionality reduction. Fig. 3, left, shows the t-SNE plot of the balanced dataset, where the red dots indicate fault events, and the blue dots indicate samples without fault events. This colouring scheme is the same for other plots unless other is specified. The t-SNE algorithm compress the data into two dimensions that is hard to interpret. Since the value of the axes dont provide any information, the axis labels have been removed for a clear, concise figure. From the mentioned figure there are some distinguishable clusters, but overall the samples with and without events overlap. If the separation of fault and non fault events had been trivial, the plot would have shown two distinct groups, with red and blue dots.

Conducting t-SNE on the unbalanced dataset yields a 2D representation displayed in Fig. 3, right. It should be noted that the result does not show any sign of clear clusters, although some separation of samples are done.

It is recommended to reduce the dimensionality of the dataset before applying t-SNE, to reduce noise and to speed up the computation. [11] PCA was used on both the balanced and unbalanced dataset. The results are displayed in Fig. 4. Comparing Fig. 4 with Fig. 3, it is clear that t-SNE did a better job on separating the samples based on the event type after doing PCA on the datasets. Due to the different nature of the two datasets, the perplexity and learning rate differ in the two cases. This is according to best practices and substantial efforts are done in finding suitable values [11].

B. k-means

Based on the recommendations on reducing the dimension of the dataset to suppress noise, PCA was applied to both the balanced and unbalanced datasets.

K-means clustering was applied on the dimensionality-reduced dataset. The algorithm was set to identify 20 clusters. Table I shows the groups with the highest fraction of faults in them. It is clear that the algorithm can to some extent

This is the accepted version of a paper published in 2020 International Conference on Smart Energy Systems and Technologies - SEST
http://dx.doi.org/10.1109/SEST48500.2020.9203294

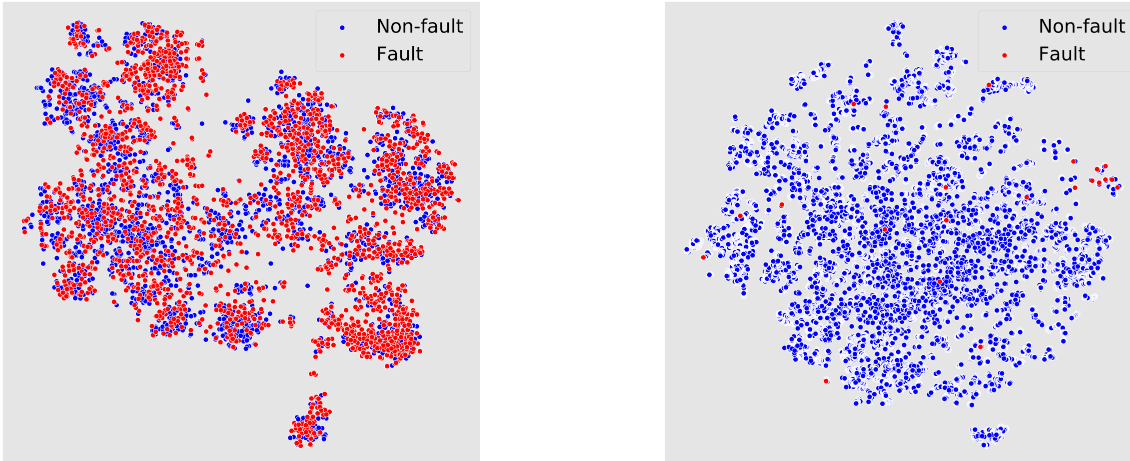

Fig. 3. Left: t-SNE reduced 2-dimensional representation of the balanced dataset.(lr: 200, p: 30, ni: 26000.) Right: t-SNE reduced 2-dimensional representation of the unbalanced dataset.(lr: 50, p: 100, ni: 60000.)

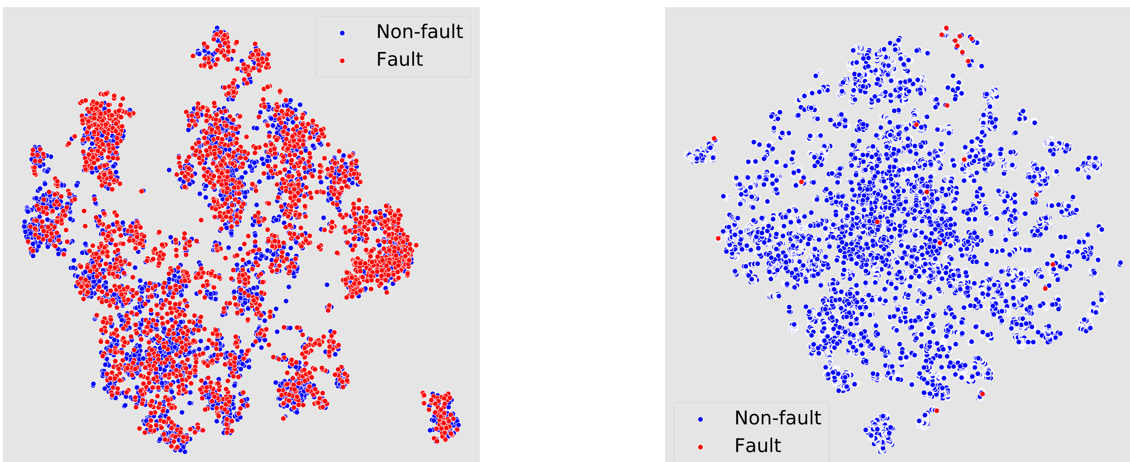

Fig. 4. Left: t-SNE reduced 2-dimensional representation of the balanced dataset with PCA reduction. (lr: 200, p: 30, ni: 26000.) Right: t-SNE reduced 2-dimensional representation of the unbalanced dataset with PCA reduction. (lr: 50, p: 100, ni: 60000.

distinguish between fault and non-fault events. In group 18, 82.5% of the samples are faults.

For visualization of the k-means algorithm, the t-SNE plot in Fig. 3 are kept. The samples are coloured according to which group k-means predict that they belong to, and plotted together with samples with fault-events represented as a black dot. This is shown in Fig. 5, left. The groups containing the highest fraction of samples with events are shown in Fig. 5, right. From this figure it is clear that meaning-full separation of groups can be done, both by t-SNE and k-means on the balanced dataset.

The unbalanced dataset was treated the same way as the balanced dataset. In Fig. 6, all gruoups, and groups with highest percentage of faults are shown. The samples with fault-events are also here represented by a black marker.

The groups containing the highest fraction of samples with events are seen in Table II. From the table and from Fig 6, it is seen that the algorithms do group many of the fault events together, but they are not distinguishable from non-fault events. For example, in Group 4, contains 7 faults but also 722 non-faults.

TABLE I
TOP FIVE CLUSTER GROUPS ON BALANCED DATASET.

Groups	Faults	Events	Faults / Events	Faults / All faults
18	132	160	0.825	0.058
17	80	99	0.808	0.035
2	109	148	0.736	0.048
14	77	109	0.706	0.034
1	205	323	0.635	0.089

This is the accepted version of a paper published in 2020 International Conference on Smart Energy Systems and Technologies - SEST
http://dx.doi.org/10.1109/SEST48500.2020.9203294

Fig. 5. Left: t-SNE on PCA reduced balanced dataset. All samples are plotted in the xy plane according to their t-sne coordinates. The samples are colored by which group k-means predict they belong to. The faults are plotted (again) as black dots.
Right: Plot of the group having the highest fraction of samples with events.

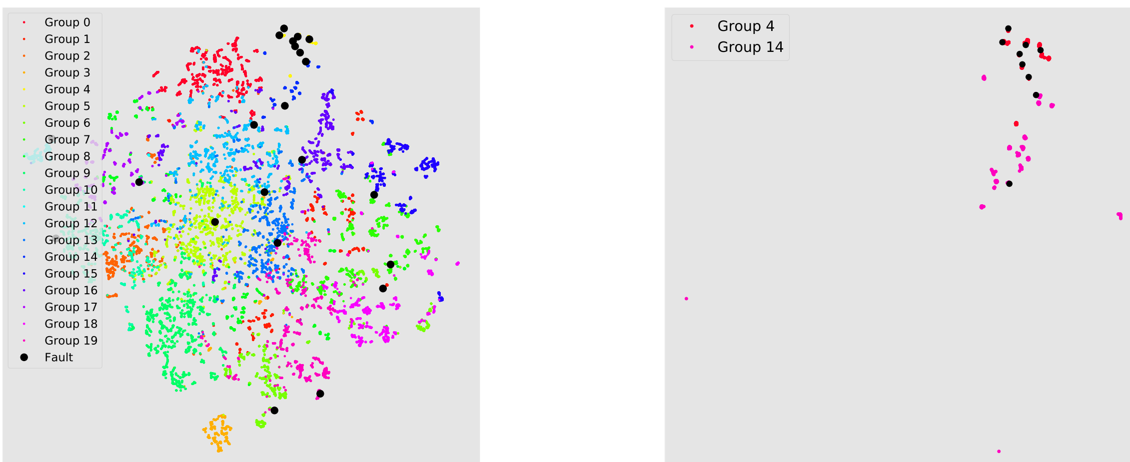

Fig. 6. Left: t-SNE on PCA reduced unbalanced dataset. All samples are plotted in the xy plane according to their t-sne coordinates. The samples are colored by which group k-means predict they belong to. The faults are plotted (again) as white dots.
Right: Plot of the group having the highest fraction of samples with events.

TABLE II
TOP FIVE CLUSTER GROUPS ON UNBALANCED DATASET.

Groups	Faults	Total	Faults / Total	Faults / All faults
4	7	729	0.010	0.318
14	2	925	0.002	0.091
19	2	3828	0.001	0.091
11	1	857	0.001	0.045
18	0	3095	0.000	0.000

V. DISCUSSION

In a balanced dataset containing samples with and without events, it is possible to extract some information to separate the different sample types, as seen from Fig. 3, left. From this figure, a clear grouping of samples with events can be seen towards the lower center-right.

When the dataset becomes unbalanced, where 0.03 % of the samples are samples with events, the information describing the event samples drowns in the noise of samples without events, as seen from Fig. 3, right.

To reduce the amount of noise, dimensionality reduction by PCA was applied to the datasets. By doing this (Fig. 4, left), the separation of the samples within the balanced dataset was better compared to the non-reduced dataset. Despite this, only small improvements of separation can be found in the reduced unbalanced dataset, see Fig. 4, right. This indicates that in a dataset with relatively similar samples, the variation among the samples is more significant than the variation between the event types when there is an unequal distribution of sample types.

The main scope of this research was to investigate prediction possibilities using unsupervised learning, applying k-means clustering to the dataset. Fig. 5, left, shows a dimensionality-

This is the accepted version of a paper published in 2020 International Conference on Smart Energy Systems and Technologies - SEST
http://dx.doi.org/10.1109/SEST48500.2020.9203294

reduced representation of the clustered dataset. The clustering algorithm is able to separate most of the same samples as the dimensionality-reducing algorithm. Considering the top tree clustering groups (isolated in Fig. 5, right), they are able to cover 10 % of the events, where 79 % of all samples within the clusters were samples with events. Clustering on the unbalanced dataset, on the other hand, is able to cover 32 % of all events in one cluster. Despite this, the grous also contain a large amount of samples without events, leaving the total fraction of samples with events in the groups to be 1 %, as detailed in Table II.

These various tests show that it is possible to isolate information from the signal when enough samples are present. It also shows that if a type of sample is undersampled, the information from these samples may drown in the noise from the majority sample type. To conquer this problem, a more directed work on pre-processing should be applied to remove most of the unnecessary noise.

The above results illustrate that problems that seem manageable in the case of event separation with balanced datasets drastically declines in performance when the categories in the dataset become unbalanced. This is very relevant for predicting events in the power grid, since there is no events in the vast majority of time series.

The presented work show that there is still some work to be done to rely solely on clustering for predicting fault events in the power system. The research is successful in automatically sort data-samples from the grid into groups with different frequency of fault-events. These findings provide interesting next steps in an automatic event-detection decision support tools for power systems operators. By automatic clustering, the overwhelming amounts of real-time sampled data can be refined into more concise datasets, that can be studied in depth. For example, the authors suggest to investigate the groups shown in Table I further. By adding more non-fault events, supervised learning can be leveraged to distinguish the faults and non-faults in the group. The authors suggest to further pursue feature engineering, other relevant data soruces (grid-topology, weather data) and various sampling techniques.

VI. CONCLUSIONAND FURTHER WORK

This work demonstrates that clustering and dimensionality-reduction techniques can be applied in when trying to predict fault-events in the power grid. The algorithms show limited predictive capability, but should be utilized in combination with e.g. supervised learning algorithms. The paper also describe how the predictive capability is reduced when treating unbalanced rather than balanced datasets, an important discovery as sequential real-time data are unbalanced.

To further develop this line of investigative techniques, there are a few avenues that could be pursued. Firstly, focus should be given to the pre-processing of the data in order to establish the *normal variation under standard operating conditions*. That is, continue to improve how to represent fault-event in the power grid, while leveraging insights from experts in the field and state of the art statistical representation of complex data. This work would give a robust statistical foundation as to the variations that should be regarded as *noise in normal operations*. Having established this, feature engineering can be performed, where parameters holding these normal variations can be given less significance and the parameters with less such variations can be given more significance. Furthermore, the authors acknowledge the rapid expansion the state of the art in machine learning, and recommend investigating other in-depth unsupervised learning algorithms.

Another possible route for increasing the predictive capabilities of the data-driven models is the inclusion of data from other sources than the PQA instruments used in this paper. It is plausible that the inclusion of weather data, data on electric consumption and system configuration could increase the predictive capabilities of these methods.

REFERENCES

[1] Y. Zhou, R. Arghandeh, and C. J. Spanos, “Partial Knowledge Data-Driven Event Detection for Power Distribution Networks,” *IEEE Transactions on Smart Grid*, 2018.

[2] K. Manivinnan, C. L. Benner, B. Don Russell, and J. A. Wischkaemper, “Automatic identification, clustering and reporting of recurrent faults in electric distribution feeders,” in *2017 19th International Conference on Intelligent System Application to Power Systems, ISAP 2017*, 2017.

[3] S. Khokhar, A. A. Mohd Zin, A. P. Memon, and A. S. Mokhtar, “A new optimal feature selection algorithm for classification of power quality disturbances using discrete wavelet transform and probabilistic neural network,” *Measurement: Journal of the International Measurement Confederation*, vol. 95, pp. 246–259, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.measurement.2016.10.013

[4] O. Zyabkina, M. Domagk, J. Meyer, and P. Schegner, “A feature-based method for automatic anomaly identification in power quality measurements,” *2018 International Conference on Probabilistic Methods Applied to Power Systems, PMAPS 2018 - Proceedings*, pp. 1–6, 2018.

[5] V. Hoffmann, K. Michalowsa, C. A. Andresen, and B. N. Torsæter, “Incipient Fault Prediction in Power Quality Monitoring,” in *25 th International Conference on Electricity Distribution APPLICATION OF DYNAMIC TRANSFORMER RATINGS TO INCREASE THE 25 th International Conference on Electricity Distribution Madrid , 3-6 June 2019*, no. June, 2019, pp. 3–6.

[6] K. W. Høiem, “Predicting Fault Events in the Norwegian Electrical Power System using Deep Learning - A Sequential Approach,” *MSc Thesis*, 2019.

[7] V. M. Santi, “Predicting faults in power grids using machine learning methods,” *MSc Thesis*, 2019.

[8] C. A. Andresen, B. N. Torsæter, H. Haugdal, and K. Uhlen, “Fault Detection and Prediction in Smart Grids,” in *9th IEEE International Workshop on Applied Measurements for Power Systems, AMPS 2018 - Proceedings*, 2018.

[9] C. J. Burges, “Dimension reduction: A guided tour,” *Foundations and Trends in Machine Learning*, vol. 2, no. 4, pp. 275–365, 2009.

[10] L. J. P. Van Der Maaten, E. O. Postma, and H. J. Van Den Herik, “Dimensionality Reduction: A Comparative Review,” *Journal of Machine Learning Research*, vol. 10, pp. 1–41, 2009.

[11] Laurens van der Maaten and G. Hinton, “Visualizing Data using t-SNE,” *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

[12] T. de Bruin, K. Verbert, and R. Babuska, “Railway Track Circuit Fault Diagnosis Using Recurrent Neural Networks,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 3, pp. 523 – 533, 2017.

[13] P. Berkhin, “A survey of clustering data mining techniques,” in *Grouping Multidimensional Data: Recent Advances in Clustering*, 2006.

[14] A. K. Jain, M. N. Murty, and P. J. Flynn, “Data clustering: A review,” in *ACM Computing Surveys*, 1999.

[15] R. Xu and D. Wunsch, “Survey of clustering algorithms,” 2005.

[16] A. K. Jain, “Data clustering: 50 years beyond K-means,” *Pattern Recognition Letters*, 2010.

[17] T. Warren Liao, “Clustering of time series data - A survey,” *Pattern Recognition*, 2005.

This is the accepted version of a paper published in 2020 International Conference on Smart Energy Systems and Technologies - SEST
http://dx.doi.org/10.1109/SEST48500.2020.9203294