Power System Event Identification based on Deep Neural Network with Information Loading

Jie Shi, Student Member, IEEE, Brandon Foggo, Member, IEEE, Nanpeng Yu, Senior Member, IEEE,

Abstract—Online power system event identification and classification is crucial to enhancing the reliability of transmission systems. In this paper, we develop a deep neural network (DNN) based approach to identify and classify power system events by leveraging real-world measurements from hundreds of phasor measurement units (PMUs) and labels from thousands of events. Two innovative designs are embedded into the baseline model built on convolutional neural networks (CNNs) to improve the event classification accuracy. First, we propose a graph signal processing based PMU sorting algorithm to improve the learning efficiency of CNNs. Second, we deploy information loading based regularization to strike the right balance between memorization and generalization for the DNN. Numerical studies results based on real-world dataset from the Eastern Interconnection of the U.S power transmission grid show that the combination of PMU based sorting and the information loading based regularization techniques help the proposed DNN approach achieve highly accurate event identification and classification results.

Index Terms—Event identification, deep neural network, graph signal processing, information loading, phasor measurement unit.

I. INTRODUCTION

DRIVEN by the need to improve the reliability of the power grid following the 2003 blackout in the North-eastern United States, phasor measurement unit (PMU) usage has experienced exponential growth. Nearly 2,000 PMUs are currently deployed in North America and over 3,000 PMUs are currently commissioned in the Chinese power grid [11]. By leveraging the fast streaming PMU data, various algorithms have been developed to enhance power system operators’ situational awareness.

To further improve system reliability, a highly accurate and automatic event detection and identification algorithm is in critical need. When power system events are correctly detected and classified in a timely manner, appropriate corrective control actions can be taken by system operators or control systems to prevent blackouts. Many power system event detection algorithms using PMU data have been developed. However, very few researchers have explored the event classification problem due to a general lack of access to large amounts of real-world streaming PMU data. By leveraging over 2 years of streaming data from 187 PMUs and over 1,000 events, this paper develops an accurate deep neural network based power system event identification and classification algorithm.

The topic of data-driven power system event detection has been studied extensively. We summarize the related literature which can be grouped into five categories. The algorithms in the first category detect power system events by performing spectral analysis such as wavelet transforms [2], [3], short-time Fourier transforms [4], graph Fourier transforms [5], and self-coherence method [6] on the PMU data. The algorithms in the second category first develop forecasts of PMU data, then an event is detected if the forecast error exceeds some thresholds [7], [8], [9]. The third category algorithms monitor the variation of spatial correlations among different PMUs via the correlation coefficient matrix [10], the sample covariance matrix [11], and the tensor sample covariance matrix [12]. A large variation in spatial correlations indicates the occurrence of a power system event. The fourth category of approaches exploits the low-rank property of PMU data during non-event periods. A significant increase in data matrix rank is treated as the sign for a power system event [13], [14]. The last group of algorithms use data mining techniques such as matrix profile [15] to detect power system events.

Although the sub-field of data-driven event detection has seen tremendous development, the subject of power system event identification and classification has not been fully explored. The insufficient research in this area is mainly due to the lack of access to large-scale labeled real-world PMU data. Most of the existing work leverages just a small amount of real-world synchrophasor data and a limited number of event labels from a restricted class of events. For example, the dataset of reference [16] only contains 32 labeled power system events. Reference [8] uses historical data of a single PMU, and reference [17] uses just 4 PMUs for its case studies.

To overcome the challenges associated with the lack of access to real-world data, some researchers have tried to create and leverage synthetic PMU data [18], [19], [20]. However, this workaround has its own drawbacks. It is extremely difficult to generate large-scale noisy streaming PMU data during power system events with time-varying spatial temporal correlations similar to that of the real-world data.

Two recent works have managed to collect a relatively large amount of synchrophasor data. Reference [21] used one-year of historical data from 44 PMUs in the Pacific Northwest.
However, the dataset only contains 57 labeled line events, which limit its use for training a general event classifier. Similarly, reference [22] used hundreds of labeled frequency events from the FNET/GridEye system to train a frequency event detection algorithm based on a deep neural network. However, the lack of labels for other event types makes it infeasible to develop a general event identification and classification model.

Equipped with two years of data from hundreds of PMUs and over one thousand event labels of different types in the Eastern Interconnection of the U.S. power transmission grid, this paper aims at developing a deep neural network based framework to identify and classify power system events in real time. To deal with high-dimensional PMU input arrays, i.e., tensors, we adopt the convolutional neural networks (CNNs) as the base model. CNNs enable sparse interaction which not only greatly reduces the memory requirements of the model, but also improves its statistical efficiency. To make parameter sharing more effective in the CNN framework, we propose an innovative graph signal processing (GSP) based PMU sorting algorithm that systematically arranges PMUs in the input tensor. To further improve the event identification and classification accuracy, we propose an information loading based regularization technique to control the amount of information compression between the input layer and the last hidden layer of the deep neural network.

The unique contributions of this work are listed as follows:

- We develop a novel deep neural network with information loading based regularization for power system event identification and classification.
- We propose a GSP based PMU sorting algorithm to make the parameter sharing scheme more effective in the proposed CNN framework.
- Our proposed deep neural network based approach achieves a high F1 score on the power system event classification task using real-world PMU data.
- The representations learned by the deep neural network are interpretable and meaningful.

The rest of the paper is organized as follows: Section II presents the problem formulation and overall framework of the proposed deep neural network based power system event identification algorithm. Section III provides the technical methods of the proposed approach. Section IV validates the proposed power system event identification framework with real-world PMU data. Section V states the conclusions.

II. PROBLEM FORMULATION AND OVERALL FRAMEWORK

In this section, we present the problem formulation and the overall framework of the proposed approach for power system event identification and classification. The power system event identification problem can be formulated as a statistical classification problem. We develop a deep neural network and train it with historical PMU data with the corresponding power system event labels. When the training process completes, the fitted model servers as an online classifier to identify different power system events using streaming PMU data.

The overall framework of the proposed approach is shown in Figure 1. The overall framework has three key modules: the neural classifier, the GSP based PMU sorting algorithm, and the information loading based regularization technique.

The raw PMU input data array to the proposed algorithm is a 3-dimensional tensor of real power ($P$), reactive power ($Q$), voltage magnitude ($|V|$), and frequency ($f$) measurements from multiple PMUs for a certain time period. The 3 dimensions of the tensor represent time, PMU ID, and $PQ|V|f$ index as illustrated in Fig. 2. For real-world power grids, the number of deployed PMUs can reach hundreds or thousands.

The GSP based PMU sorting module takes the raw tensor and outputs a sorted PMU tensor, which is more compatible with the parameter sharing scheme of the neural classifier. In the sorted tensor, PMUs with highly correlated measurements are arranged to be closer to each other. The technical methods used in the GSP based PMU sorting module is presented in Section III-A.

The CNN based classifier module first takes the sorted PMU tensors as inputs and leverage convolution filters in successive layers to transform the inputs into interpretable hidden representations. The type of power system event of the corresponding time window is then identified by the estimation layer. The overall neural network design is described in Section III-C.

During the training process, the information loading based regularization module will estimate the mutual information between the input tensors and the hidden representations and use it to augment the typical cross-entropy loss function. The information loading technique is based on the information bottleneck theory [23] and our recent theoretical results on new bounds of information losses in neural classifiers [24]. The theoretical foundation of information losses and technical methods of information loading based regularization will be presented in Section III-B.

III. TECHNICAL METHODS

The technical methods of the proposed power system event identification approach are presented in this section. The GSP
A. Graph Signal Processing Based PMU Sorting

In this subsection, we propose a sorting algorithm based on GSP to systematically arrange PMUs in the \( PQ \mid f \) tensor. In order to leverage the powerful convolutional operation in deep neural network design, we strategically place highly correlated PMUs close to each other. Again, the goal is to determine a useful ordering of PMUs in the input tensor. Let \( N \) denote the total number of PMUs and \( d_i \) be the index of the \( i \)th PMU in the ordered input tensor.

The problem of finding \( d = \{ d_1, \cdots, d_N \} \) such that similar PMUs are arranged closer together can be formulated as the following optimization problem:

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} W_{ij} (d_i - d_j)^2 \\
\text{subject to} & \quad d^T d = 1 \\
& \quad d^T 1 = 0
\end{align*}
\]  

(1)

(2)

(3)

where \( W_{ij} \) represents the absolute value of the correlation coefficient between PMUs \( i \) and \( j \)'s measurements.

The objective function (1) represents a correlation-weighted sum of PMU index squared distances. Constraints (2) and (3) ensure that the index of the PMUs in the ordered tensor are centered around the origin and not placed at one single point. This optimization problem is nonlinear and non-convex due to the quadratic constraint. However, the problem can be efficiently solved with GSP based techniques.

We can envision that all PMUs (vertices) are connected to each other in a complete graph. Let \( G = (\mathcal{V}, \mathcal{E}) \) be a complete graph, where \( \mathcal{V} = \{ v_1, \cdots, v_N \} \) is a set of vertices representing the PMUs within a power system. \( \mathcal{E} = \{ e_{ij} | i \neq j, i, j = 1, \cdots, N \} \) is a set of edges where \( e_{ij} \) stores the Pearson correlation coefficient between the \( i \)th and \( j \)th PMU’s measurements. Define \( W \) as a weight matrix with diagonal elements being zeros. Its non-diagonal element \( W_{ij} \) are set as \( |e_{ij}| \). Let \( L = D - W \) be the graph Laplacian, where the degree matrix \( D \) is a diagonal matrix with \( D_{ii} = \sum_{j=1}^{N} W_{ij} \). Then we have the following relationship (23):

\[
\begin{align*}
d^T L d &= \sum_{i=1}^{N} \sum_{j=1}^{N} W_{ij} (d_i^2 - d_i d_j) \\
&= \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} W_{ij} (d_i^2 - d_i d_j + d_j^2 - d_j d_i)^2 \\
&= \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} W_{ij} (d_i - d_j)^2
\end{align*}
\]

(4)

By substituting (4) and (2) into (1), the objective function (1) can be converted to the Rayleigh quotient \( d^T L d / d^T d \). The converted unconstrained objective function has a minimum value equal to the smallest eigenvalue of the graph Laplacian \( L \), which is 0. The optimal solution to the unconstrained objective function is the eigenvector of \( L \) that corresponds to the smallest eigenvalue, which is \( 1 / \sqrt{N} \). However, (3) rules out this solution by constraining \( d \) to be orthogonal to \( 1 / \sqrt{N} \). Therefore, the solution of constrained optimization problem (1)-(3) is the eigenvector corresponding to \( L \)'s second smallest eigenvalue. Thus, the PMUs' measurements should be sorted according to the optimal solution \( d \) in an ascending order. Algorithm 1 summarizes the procedures of the proposed GSP based PMU sorting approach.

B. Information Loading based Regularization

The motivation and algorithm for the information load based regularization technique are presented in this subsection. Information loading is a regularization technique first proposed in our previous work [26]. It is motivated by the information bottleneck theory [23] and recent theoretical results about information losses of neural classifiers [24]. The information loading based regularization technique controls the amount of information compression between the input layer and the last hidden layer of a deep neural network. First, we briefly review the theory of information losses and then present the information loading algorithm.

1) Information Losses: A general classification model can be represented as a Markov chain as shown in Fig. 3. \( Y \) and \( \hat{Y} \) denote the real class label and the estimated label. \( X \) denotes the input features. \( Z \) denotes the learned representation (in our case, the last hidden layer of a deep neural network). \( y, \hat{y}, x, \) and \( z \) are instances of the random variables \( Y, \hat{Y}, X, \) and \( Z \), respectively. The classifier consists of two parts: an encoder that models \( P_{Z \mid X} \) and an estimator that models \( P_{Y \mid Z} \).

To achieve better classification accuracy, we want to learn a representation \( Z \) which has a high mutual information \( I(Y; Z) \) with the class label. But in practice, \( I(Y; Z) \) cannot be estimated from training data alone. Instead, we must resort to using an estimate, denoted \( \hat{I}(Y; Z) \), which has the following form:

\[
\hat{I}(Y; Z) = \mathbb{E} \left[ \log_2 \frac{P_{ZY}}{P_{Z}P_{Y}} \right]
\]

(5)

Algorithm 1: GSP based PMU sorting algorithm

1. Obtain the Pearson correlation coefficients between PMUs;
2. Construct weight matrix \( W \) and Laplacian graph \( L \);
3. Take eigendecomposition of \( L \);
4. Sort PMUs according to the eigenvector corresponding to the second smallest eigenvalue of \( L \);

Fig. 3. Flowchart of a general classification model.
where \( \hat{P}_{ZY} \), \( \hat{P}_Z \), and \( \hat{P}_Y \) denote the estimates of \( P_{ZY} \), \( P_Z \), and \( P_Y \) based on the training samples.

Let \( Z^* \) and \( \tilde{Z} \) be the representations that, respectively, maximize \( I(Y; Z) \) and \( I(X; Z) \) with the constraint \( I(X; Z) = C \), where \( C \) is a constant that defines the complexity of encoder. In other words, \( Z^* \) is produced by the optimal encoder when we have perfect knowledge of \( P_{XY} \). \( \tilde{Z} \) is produced by the optimal encoder when we have partial knowledge of \( P_{XY} \).

Now, we can define the term “information losses”, which is strongly related to ‘minimal classification error’ [26]:

\[
I_{loss} = |I(Y; Z^*) - I(Y; \tilde{Z})|
\]

The upper bound of the “information losses” is given by [24]:

\[
I_{loss} \leq 2 \left( \delta_{\tilde{Z}} I(X; \tilde{Z}) + h_2(\delta_{\tilde{Z}}) \right) + \epsilon
\]

(7)

where \( h_2(\cdot) \) denotes the binary entropy function. \( \delta_{\tilde{Z}} \) is the conditional total variation of \( \hat{P}_{Y|X} \) from \( P_{Y|X} \):

\[
\delta_{\tilde{Z}} = \frac{1}{2} \mathbb{E}_{P_X} \left[ \sum_{y \in Y} \left| P_{Y|X}(y|x) - \hat{P}_{Y|X}(y|x) \right| \right]
\]

(8)

It is worth noting that \( \delta_{\tilde{Z}} \) has an upper bound that only depends on the given training data [27].

The upper bound of information losses (7) has been shown to be reasonably tight across many testing datasets [24]. It will serve as the cornerstone in the information loading based regularization technique.

2) Information Loading: Based on the theoretical results of information losses, we can visualize the mutual information bounds for typical datasets. Suppose \( I(Y; Z^*) \) is given, we can draw the lower bound of \( I(Y; \tilde{Z}) \) based on (6) and (7). It has been shown that this lower bound behaves differently depending on the entropy level of the input feature space [24] as illustrated in Fig. 4.

For low entropy features, the lower bound keeps increasing as the complexity of the encoder \( I(X; \tilde{Z}) \) increases. In this case, we can safely boost the mutual information between the input and the hidden representation \( I(X; \tilde{Z}) \) to achieve higher \( I(Y; \tilde{Z}) \) and classification accuracy. However, we need to be careful about adjusting \( I(X; \tilde{Z}) \) for high entropy input data. If the representation becomes too complex, the performance of the classifier will deteriorate.

Previous research [23] shows that during the training process, deep neural networks often compress information between the input layer and the last hidden layer, i.e., reducing \( I(X; \tilde{Z}) \). Thus, we can penalize the compression of information by augmenting the typically cross-entropy loss function of classification model with \(-\beta I(X; \tilde{Z})\). This will help \( I(Y; \tilde{Z}) \) approach its peak.

The selection of hyperparameter \( \beta \) depends on the entropy of the input feature space. As explained before, for a low entropy feature space, \( \beta \) could be selected to be a larger value (in accordance with the left hand side of Fig. 4). However, for a high entropy feature space, a very large \( \beta \) may lead to performance degradation (in accordance with the right hand side of Fig. 4).

C. Neural Network Design for System Event Identification

In this subsection, we present the neural network design for power system event identification using PMU data. The overall neural network architecture of the proposed solution is shown in Fig. 5. It is composed of three main modules: the classifier model, the mutual information estimator (MIE), and the information loaded loss function. The design for each of the three modules is described in detail below.

1) The classifier model: The classifier is an essential module of the overall design. Both the MIE and the information loaded loss function modules are built to facilitate the training of the classifier. Once trained, the classifier will serve as a standalone unit to identify power system events based on streaming PMU data.

The proposed classifier module contains two components: the encoder and the estimator. The encoder transforms the input features into representations. The representations are then fed into the estimator, which produces the final power system event identification results. The input \( PQ[f] \) tensors are arranged based on the GSP based PMU sorting algorithm described in Section III.A. Since the input \( PQ[f] \) tensors have structures similar to the images, we decided to adopt a widely-used deep convolutional neural network (CNN), ResNet-50 [28] as a key building block of the encoder. ResNet and its extensions have achieved great success across various applications, making them one of the most popular deep CNN families in the machine learning community. In this work, the encoder is built by connecting ResNet-50 (excluding the output layer) to a dense layer of 10 neurons. The estimator is designed as one dense layer with the softmax activation function.

2) Mutual Information Estimator: The mutual information estimator is built to provide an estimate of mutual information
between the input feature \( X \) and the representation \( Z \). As discussed in Section III-B, we need to tune \( I(X; Z) \) to improve the classification accuracy. The MIE estimates \( \hat{I}(X; Z) \), which will be fed into the information loaded loss function module.

The proposed mutual information estimator is inspired by MINE-f introduced in [29]. By definition, \( I(X; Z) \) can be written in the KL-divergence form:

\[
I(X; Z) = D_{KL}(P_{XZ} \Vert P_X \otimes P_Z) \tag{9}
\]

where \( P_{XZ} \otimes P_Z \) denotes the product of the marginal distributions \( P_X \) and \( P_Z \). The \( f \)-divergence representation developed by [30], [31] provides a lower bound of (9):

\[
D_{KL}(P_{XZ} \Vert P_X \otimes P_Z) = \sup_{g \in \mathcal{G}} \mathbb{E}_{P_{XZ}}[g] - \mathbb{E}_{P_X} \mathbb{E}_{P_Z}[e^{g-1}] \tag{10}
\]

where \( \mathcal{G} \) denotes the set of all possible mappings. In this study, we parameterize \( g \) using a deep neural network \( g_\theta \). As shown in Fig. 5, \( g_\theta \) is composed of two parts: a compression net and a feed-forward neural network (FNN). The compression net is used to compress the input \( PQ|V|f \) tensor into a low-dimensional representation. Specifically, we employ a lightweight deep CNN called MobileNetV2 [32] as the compression net. The FNN has two layers of neurons with dimensions of 200 and 1.

Based on (9) and (10), we can build an estimate of \( I(X; Z) \) by drawing samples from the input feature space and the representation as follows:

\[
\hat{I}(X; Z) = \frac{1}{B_I} \sum_{i=1}^{B_I} g_\theta(x_{1,i}, z_{1,i}) + \frac{1}{B_I} \sum_{i=1}^{B_I} e^{g_\theta(x_{1,i}, z_{2,i})-1} \tag{11}
\]

where \( B_I \) denotes the training batch size. \( x_{1,i} \) and \( z_{1,i} \) are samples drawn from the joint distribution \( P_{XZ} \), \( z_{2,i} \) is a sample drawn from the marginal distribution \( P_Z \). During the training session, \( z_1 = [z_{1,1}, \cdots, z_{1,B_I}] \) and \( z_2 = [z_{2,1}, \cdots, z_{2,B_I}] \) are obtained by feeding two independent input batches \( x_1 = [x_{1,1}, \cdots, x_{1,B_I}] \) and \( x_2 = [x_{2,1}, \cdots, x_{2,B_I}] \) into the encoder (See Fig. 5).

3) Information loaded loss function: The loss function \( L_T \) of the proposed power system event identification model contains two components.

\[
L_T = L_{CE} - \beta \hat{I}(X; Z) \tag{12}
\]

The first component \( L_{CE} \) is the typical cross-entropy loss function, which is often used for training classifiers. \( L_{CE} = -\sum_{i=1}^{B_I} y_i \cdot \log \hat{y}_i \), where \( y_i \) and \( \hat{y}_i \) denote the true label and the estimated label of input sample \( x_{1,i} \), respectively.

The second component of the information loaded loss function, \(-\beta \hat{I}(X; Z)\), penalizes the compression of information between the input feature and the representation of the deep neural network. The hyperparameter \( \beta \) regularizes the mutual information \( I(X; Z) \) to boost the performance of the neural classifier. The Adam optimizer [33] is adopted to minimize the information loaded loss function during training.

IV. NUMERICAL STUDY

In this section, we validate our proposed power system event identification algorithm using a large-scale real-world PMU dataset. First, we briefly discuss the data source and the data preprocessing steps. Then, we present the GSP based PMU sorting results for our PMU dataset. Next, we explain how to address the class imbalance issue with data augmentation. Finally, we carry out an ablation study to quantify the benefits of the GSP based PMU sorting technique and the information loaded based regulation method in improving the power system event identification performance.

A. Data Source

The dataset is comprised of two years of PMU data from the Eastern Interconnection of the continental U.S. power transmission grid prepared by the Pacific Northwest National Laboratory. The raw data from 187 PMUs include measurements of frequency and positive sequence voltage/current magnitudes and phasor angles. The reporting frequency of the PMU data is 30 Hz. The specific location of the PMUs are concealed by the original data provider per the non-disclosure agreement. We convert the raw readings from the PMUs into the corresponding positive sequence real power (\( P \)), reactive power (\( Q \)), voltage magnitude (\( |V| \)), and frequency (\( f \)) data arrays, i.e., the \( PQ|V|f \) tensors. The raw data includes a total 1147 labeled power system events and 120 non-events. The time span of each labeled event data sample is 20 seconds with the event starting time in the middle of the window. 8 PMUs are removed from the analysis due to prevalent bad data. Thus, the dimensionality of the input \( PQ|V|f \) tensors is \( [600, 179, 4] \), where \( 600 = 20 \times 30 \) corresponds to the number of time stamps in the 20-second window and 179 is the number of valid PMUs.

Four types of labels are provided for the \( PQ|V|f \) tensors. There are 120 Non-events, 825 Line events, 84 Generator events, and 118 Oscillation events. The \( PQ|V|f \) tensor of a sample generator tripping event is depicted in Fig. 6. The Non-event class corresponds to time periods without any observable
power system events. The Line event class includes line tripping events and line fault events. The Generator-event class includes generator tripping events. The Oscillation event class consists of power oscillation events. Note that our dataset is severely imbalanced due to the relatively large number of line events. This issue will be addressed by the data augmentation procedure in subsection IV.C.

B. GSP Based Sorting Result

We calculate the Pearson correlation coefficient for each pair of PMUs. The entries of weight matrix $W$, which quantify the correlation between PMUs, are derived from the absolute values of the corresponding correlation coefficients. Fig. 7 compares the weight matrices of the original PMU sequence and the sorted PMU sequence. As shown in the figure, highly correlated PMUs are placed much closer to each other after the GSP based PMU sorting. It will be shown in IV-D the GSP based PMU sorting technique greatly facilitates the kernel learning in CNN and significantly improves the event classification performance.

C. Data Augmentation

Two issues exist in our dataset. First, the original $P|V|f$ tensors are severely imbalanced due to the relatively large number of line events. The class imbalance can result in over-classification of the majority group due to biased prior distribution [34]. Second, the event starting time for each event sample is always located in the middle of the time window. In other words, the event signatures consistently appear in the center of the corresponding $P|V|f$ tensors, leading to a biased distribution of event timing. Data augmentation is thereby introduced to address these two issues.

The core idea of the proposed data augmentation technique is straightforward. We sample sub-tensors from the original $P|V|f$ tensors independently and uniformly. Fig. 8 illustrates that a 12-second sub-tensor is sampled uniformly from a given $P|V|f$ tensor. Hereafter we will refer to the sampled sub-tensors $P|V|f$ as snapshots. It is worth noting that the time range of $P|V|f$ snapshots should be large enough to capture the low-frequency power oscillations according to Nyquist–Shannon sampling theorem. In this study, we set this time range to be 12 seconds.

To address the two issues associated with imbalanced dataset, we sample 6, 1, 9, and 6 $P|V|f$ snapshots independently and uniformly from each $P|V|f$ tensor in Non-event, Line-event, Generator-event, and Oscillation-event category, respectively. This procedure creates a relatively balanced dataset as shown in Table I. Meanwhile, the event starting times are no longer fixed in the middle of the input tensors resulting in increased diversity of the event timing.

D. Classification Performance

We evaluate the performance of the proposed deep neural network based power system event classification algorithm by quantifying its classification accuracy and F1 score on the real-world PMU dataset. We perform an ablation study to tease apart which component(s) of the proposed algorithm are most important for its success. To achieve this goal, we evaluate four methods. The first method directly employs a powerful CNN
We adopt z-score scaling on learning rate of the Adam optimizer is selected to be 0.001. The training epochs is 200. The size of the training batch is 16. The training session are provided as follows. The total number of train the neural network for 10 rounds. In each round, 9 subsets are used for training and the other subset is used for validation. This is called 10-fold cross-validation. The settings of each training session are provided as follows. The total number of training epochs is 200. The size of the training batch is 16. The learning rate of the Adam optimizer is selected to be 0.001. We adopt z-score scaling on \( P, Q, |V|, \) and \( f \) of each PMU in the input \( PQ|V|f \) snapshots.

The average F1 scores of the power system event classification results of the baseline+GSP+info method under the cross-validation setup is reported in Table III. As shown in the table, \( \beta = 0.1 \) achieves the best result in the cross-validation setup. Thus, the hyperparameter \( \beta \) is selected to be 0.1. The average accuracy over 10 rounds of four different methods as the training session proceeds are reported in Fig. 9a. Clearly, our proposed method baseline+GSP+info performs the best in terms of validation accuracy. The results show that the combination of GSP based PMU sorting and the information loading based regularization is capable of significantly boosting the performance of the baseline model. It is worth noting that there is no significant accuracy difference between baseline+info and baseline. This is because it is extremely difficult to learn the kernel in the convolutional layers with a random PMU sequence.

After the cross-validation is completed, we train the neural networks with the full training dataset and evaluate their performance on the testing dataset. Specifically, we repeat training and testing for 10 times with different initial neural network weights. The average testing accuracy with respect to training epoch for each method is shown in Fig. 9b. The average testing F1 scores for each power system event class is reported in Table III. The testing results show that baseline+GSP+info performs the best in terms of both classification accuracy and F1 scores for all event categories. Compared to the baseline model, the combination of GSP based PMU sorting and information loading based regularization work synergistically to boost F1 scores for non-event, line-event, generator event, and oscillation event by 75%, 10%, 9%, and 374%, respectively. The most dramatic performance improvement can be observed for non-events, line-events, and oscillation events when GSP based PMU sorting method is applied. The information loading technique is more effective in improving the event classification results for the generator events.

### E. Representation Learning Results

The performance of a classifier is primarily dependent on the quality of representations produced by its encoder. The intrinsic goal of GSP based PMU sorting and information loading based regularization is to help encoders learn better representations. Furthermore, representations with higher quality provide better interpretability for the corresponding deep neural network. In this subsection, we visualize and compare the representations learned by different methods.

#### 1. Representation Learning Results

Direct visualization of representations is difficult due to the high dimensionality of the encoders’ outputs. To address this issue, researchers typically adopt linear dimensionality reduction methods such as principal component analysis (PCA) to reduce the dimensionality of the representations to 2 dimensions. Specifically, the principal components are determined through eigendecomposition of sample covariance matrix derived from the representation samples. Then, the original representation data points are projected onto the first two principal components, creating a 2D data array.
It is worth noting that linear dimension reduction techniques are preferred over nonlinear ones to visualize hidden representations. This is because the hidden representation layer is often followed immediately by a fully connected layer activated by the softmax function. The softmax regression, which serves as the estimator, is essentially a linear separation model. In order for the estimator to achieve great classification performance, the representations should be almost linear separable. Hence, a linear projection such as PCA is usually introduced to visualize the learned representations of the deep neural networks.

The dimension reduced hidden representations produced by different methods are shown in Fig. 10. By comparing Fig. 10a and 10b with Fig. 10c and 10d, we observe that the GSP based PMU sorting significantly improves the encoders’ ability in separating oscillation events from non-events. Meanwhile, by comparing the sub-figures on the left hand side with that on the right hand side, we see a higher separation level is gained from the information loading based regularization. Representations with this abundant separation between classes greatly simplifies the corresponding estimator’s classification task, which explains the excellent performance of Baseline+GSP+info.

V. CONCLUSION

This paper proposes to identify and classify power system events with a deep neural network using streaming PMU dataset. The proposed framework include three key components: the neural classifier, the GSP based PMU sorting method, and the information loading based regularization. The neural classifier consists of a CNN based encoder and an estimator represented by a dense layer of neurons. The GSP based PMU sorting method places highly correlated PMUs closer to each other, which makes parameter sharing more effective in the CNN based encoder. The information loading based regularization further improves the generalization of the classifier by tuning the mutual information between the input features and the representation. Testing results on large-scale PMU datasets from the Eastern Interconnection of the U.S. transmission grid demonstrate that the proposed approach achieves high accuracy in identifying power system events.

ACKNOWLEDGMENT

This material is based upon work supported by the Department of Energy under Award Number DE-OE0000916.

REFERENCES

[1] A. G. Phadke and T. Bi, “Phasor measurement units, WAMS, and their applications in protection and control of power systems,” Journal of Modern Power Systems and Clean Energy, vol. 6, no. 4, pp. 619–629, Jul. 2018.
[2] D.-I. Kim, T. Y. Chun, S.-H. Yoon, G. Lee, and Y.-J. Shin, “Wavelet-based event detection method using PMU data,” IEEE Transactions on Smart Grid, vol. 8, no. 3, pp. 1154–1162, Oct. 2015.
[3] S. S. Negi, N. Kishor, K. Uhlen, and R. Negi, “Event detection and its signal characterization in PMU data stream,” IEEE Transactions on Industrial Informatics, vol. 13, no. 6, pp. 3108–3118, Jul. 2017.
[4] S.-W. Sohn, A. J. Allen, S. Kulkarni, W. M. Grady, and S. Santoso, “Event detection method for the PMUs synchrophasor data,” in IEEE Power Electronics and Machines in Wind Applications, Jul. 2012.
[5] J. Shi, B. Foggo, X. Kong, Y. Cheng, N. Yu, and K. Yamashita, “Online event detection in synchrophasor with graph signal processing,” in IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), Nov. 2020.
[6] N. Zhou and J. Dagle, “Initial results in using a self-coherence method for detecting sustained oscillations,” IEEE Transactions on Power Systems, vol. 30, no. 1, pp. 522–530, 2015.
[7] Y. Zhou, R. Arghandehe, H. Zou, and C. J. Spanos, “Nonparametric event detection in multiple time series for power distribution networks,” IEEE Transactions on Industrial Electronics, vol. 66, no. 2, pp. 1619–1628, Jun. 2018.
[8] C. Hannon, D. Deka, D. Jin, M. Vufray, and A. Y. Lokhov, “Real-time anomaly detection and classification in streaming PMU data,” arXiv preprint arXiv:1911.08696, 2019.
[9] L. Xie, Y. Chen, and P. Kumar, “Dimensionality reduction of synchrophasor data for early event detection: Linearized analysis,” IEEE Transactions on Power Systems, vol. 29, no. 6, pp. 2784–2794, Nov. 2014.
[10] J. Wu, J. Xiong, P. Shil, and Y. Shi, “Real-time anomaly detection in wide area monitoring of smart grids,” in IEEE/ACM International Conference on Computer-Aided Design (ICCAD), Nov. 2014, pp. 197–204.
[11] Z. Ling, R. C. Qiu, X. He, and L. Chu, “A new approach of exploiting self-adjoint matrix polynomials of large random matrices for anomaly detection and fault location,” IEEE Transactions on Big Data, May 2019.
[12] X. Shi and R. Qiu, “Dimensionality increment of PMU data for anomaly detection in low observability power systems,” arXiv preprint arXiv:1910.08696, 2019.
[13] P. Gao, M. Wang, S. G. Ghiopec, J. H. Chow, B. Fardanesh, and G. Stefopoulos, “Missing data recovery by exploiting low-dimensionality in power system synchrophasor measurements,” IEEE Transactions on Power Systems, vol. 31, no. 2, pp. 1006–1013, Apr. 2015.
[14] Y. Hao, M. Wang, J. H. Chow, E. Farantatos, and M. Patel, “Modelless data quality improvement of streaming synchrophasor measurements by exploiting the low-rank Hankel structure,” IEEE Transactions on Power Systems, vol. 33, no. 6, pp. 6966–6977, Jun. 2018.
[15] J. Shi, N. Yu, E. Keogh, H. K. Chen, and K. Yamashita, “Discovering and labeling power system events in synchrophasor data with matrix profile,” in 2019 IEEE Sustainable Power and Energy Conference (iSPEC). IEEE, Nov. 2019, pp. 1827–1832.
[16] W. Li, M. Wang, and J. H. Chow, “Real-time event identification through low-dimensional subspace characterization of high-dimensional synchrophasor data,” IEEE Transactions on Power Systems, vol. 33, no. 5, pp. 4937–4947, Jan. 2018.
[17] O. P. Dahal and S. M. Brahma, “Preliminary work to classify the disturbance events recorded by phasor measurement units,” in IEEE Power and Energy Society General Meeting, Jul. 2012, pp. 1–8.
[18] S. Brahma, R. Kavasseri, H. Cao, N. Chaudhuri, T. Alexopoulos, and Y. Cui, “Real-time identification of dynamic events in power systems using PMU data, and potential applications—models, promises, and challenges,” IEEE transactions on Power Delivery, vol. 32, no. 1, pp. 294–301, Jul. 2016.
[19] H. Li, Y. Wang, E. Farantatos, and M. Patel, “An unsupervised learning framework for event detection, type identification and localization using PMUs without any historical labels,” in IEEE Power & Energy Society General Meeting (PESGM), Aug. 2019, pp. 1–5.
[20] ———, “A hybrid machine learning framework for enhancing PMU-based event identification with limited labels,” in International Conference on Smart Grid Synchronized Measurements and Analytics (SGSMA), May 2019, pp. 1–8.
[21] D. Nguyen, R. Barella, S. A. Wallace, X. Zhao, and X. Liang, “Smart grid line event classification using supervised learning over PMU data streams,” in International Green and Sustainable Computing Conference (IGSC), Dec. 2015, pp. 1–8.
[22] W. Wang, H. Yin, C. Chen, A. Till, W. Yao, X. Deng, and Y. Liu, “Frequency disturbance event detection based on synchrophasors and deep learning,” IEEE Transactions on Smart Grid, vol. 11, no. 4, pp. 3593–3605, Feb. 2020.
[23] R. Shwartz-Ziv and N. Tishby, “Opening the black box of deep neural networks via information,” arXiv preprint arXiv:1703.00810, Mar. 2017.
[24] B. Foggo, N. Yu, J. Shi, and Y. Gao, “Information losses in neural classifiers from sampling,” IEEE Transactions on Neural Networks and Learning Systems, Dec. 2019.
[25] L. Stankovic and E. Sejdic, Vertex-Frequency Analysis of Graph Signals. Springer, 2019.
[26] B. Foggo and N. Yu, “Improving supervised phase identification through the theory of information losses,” IEEE Transactions on Smart Grid, vol. 11, no. 3, pp. 2337–2346, Nov. 2019.
[27] ——, “On the maximum mutual information capacity of neural architectures,” arXiv preprint arXiv:2006.06037, 2020.

[28] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2016.

[29] M. I. Belghazi, A. Baratin, S. Rajeshwar, S. Ozair, Y. Bengio, A. Courville, and D. Hjelm, “Mutual information neural estimation,” in International Conference on Machine Learning, Jul. 2018, pp. 531–540.

[30] X. Nguyen, M. J. Wainwright, and M. I. Jordan, “Estimating divergence functionals and the likelihood ratio by convex risk minimization,” IEEE Transactions on Information Theory, vol. 56, no. 11, pp. 5847–5861, Nov. 2010.

[31] S. Nowozin, B. Cseke, and R. Tomioka, “f-GAN: Training generative neural samplers using variational divergence minimization,” in Advances in Neural Information Processing Systems 29, 2016, pp. 271–279.

[32] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “MobileNetV2: Inverted residuals and linear bottlenecks,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2018.

[33] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, Dec. 2014.

[34] J. M. Johnson and T. M. Khoshgoftaar, “Survey on deep learning with class imbalance,” Journal of Big Data, vol. 6, no. 1, p. 27, Dec. 2019.