Weakly Supervised Event Classification Using Imperfect Real-world PMU Data with Scarce Labels

Yunchuan Liu and Lei Yang
Department of Computer Science and Engineering, University of Nevada, Reno, NV, 89557

***Abstract*—This paper studies event classification using imperfect real-world phasor measurement unit (PMU) data with scarce event types (labels). By investigating the real-world PMU data, it is observed that most real-world PMU data's event type is unknown, which makes it challenging to directly use such dataset to build event classifiers as existing classification techniques require high-quality training data with known event type (i.e., label). To address this challenge, a weakly supervised learning based event classification approach is developed, which can use noisy and low-quality PMU data for the training. First, data quality issues are fixed using data preprocessing techniques and then event features are constructed from the PMU data. Using these features, a series of labeling functions are learnt to generate initial estimates of the labels of large amounts of unlabeled PMU data. As the labeling functions are learnt using the same data with scarce labels, the label estimates from the labeling functions can be correlated, noisy, and bias. To enhance these initial estimates, a generative model is developed to characterize the dependencies among the estimated labels, based on which better labels are obtained for training event classifiers. Numerical experiments using the real-world dataset from the Western Interconnection of the U.S. power transmission grid show that the proposed weakly supervised event classifier trained using the dataset with only 5% labeled data can achieve 78.4% classification accuracy.**

***Index Terms*—Phasor Measurement Units (PMUs), Event Classification, Weakly Supervised Learning**

I. Introduction

A. Motivation

Thanks to the wide deployment of phasor measurement units (PMUs), large amounts of data are being collected with high sampling rates (e.g., 30 or 60 samples per second in the U.S.), which provides golden opportunities to achieve high level of situational awareness (e.g., real-time event detection and classification). However, the collected data are often of low quality, e.g., the real-world PMU data are noisy and contain bad data, dropouts, and timestamp errors. Worse still, most of real-world PMU data's event types are unknown (unlabeled). Recently, we have opportunities to analyze large amounts of real-world PMU data provided by DOE's Big Data Analysis of Synchrophasor Data program [1], where it is observed that no event type (i.e., label) information is available in the PMU data from the Texas Interconnection of continental U.S. transmission grid. As the efficacy of existing techniques (e.g., Convolutional Neural Network [2]) for training event classifiers depends on high-quality PMU data with sufficient known event types (i.e., labels), it requires significant efforts from domain experts to maintain the event logs (i.e., to label those data) and even hand-label the event types, which can be prohibitively costly or impractical. One fundamental open question is "*How can we train a good event classification model using imperfect real-world PMU data with scarce labels in a less expensive and automated way?*"

This material is based upon work supported by the Department of Energy under Award Number DE-OE0000911. This work was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

B. Related Works

To address the challenge of insufficient labeled PMU data, some studies leverage synthetic data by simulation [3] or Generative Adversarial Neural Networks [4] to train event classification models. Although these approaches can increase the number of labeled data for training event classifiers, the event and grid characteristics hidden in the real-world PMU data can hardly be represented by synthetic data. Thus, the generalization of event classifiers trained using synthetic data can be poor.

In order to leverage the large amount of unlabeled PMU data, different unsupervised learning based labeling approaches have been proposed, e.g., clustering techniques [5], low-rank techniques [6], PCA based techniques [7], [8], and convolutive dictionary based techniques [9]. For unsupervised learning, one key challenge is how to determine the correct number of clusters (i.e., the number of event types). As no labeled PMU data are utilized, it is challenging for the unsupervised learning approaches to accurately capture the features of different types of events. Thus, it is highly possible that different types of events may be identified as the same type or normal operations could be considered as events, which would introduce modeling error in the developed event classifiers.

Different from unsupervised learning, semi-supervised learning (SSL) based methods (e.g., self-training [10], [11], hidden structure semi-supervised machine [12], and adversarial SSL [13]) have been proposed, which leverage both labeled and unlabeled data. The idea of SSL is to use limited labeled data to guide the labeling process for the unlabeled data, which can be used for training event classifiers. However, these SSL based methods rely on well-developed and non-customizable models to label the unlabeled data, and the performance of these models largely depends on the amount of available labeled data. With scarce labeled data, the performance of SSL based methods could be poor. Moreover, the SSL based methods cannot easily incorporate the domain knowledge on different types of events, which can provide useful information for event classification.

To address these challenges, a weakly supervised learning for event classification is proposed, which can use noisy and low-quality data for the training [14]. Recently, Snorkel [15], an open-source system for quickly assembling training data through weak supervision, has been developed; Snorkel employs the central principles of the data programming paradigm [16], in which developers create labeling functions to label the data and employ supervised learning techniques to assess the accuracy of these labeling functions. Using weakly supervised learning, the domain knowledge can easily be incorporated in the labeling functions, which can enable the use of low-quality unlabeled PMU data to create high-quality event classification models.

C. Main Contributions

This paper develops a weakly supervised learning based event classification approach that can train good event classifiers using imperfect real-world PMU data with scarce labels. The key idea is to estimate the labels of large amounts of unlabeled PMU data for training event classifiers. Specifically, a series of labeling functions based on the knowledge of different types of events is first learnt, which can generate initial estimates of the labels. As these labeling functions are learnt using the same dataset with scarce labels, the estimated labels are often correlated and the estimates can be noisy and biased. Directly using such estimates cannot generate good event classifiers. To enhance the accuracy of the estimated labels, a generative model is developed to characterize the correlations among the estimated labels, based on which the estimated labels from labeling functions are combined in a probabilistic manner to generate the "true" labels. Then, the refined labels from the generative model are used to train event classifiers. It is worth noting that using the proposed weakly supervised learning approach, less efforts are required from domain experts to maintain the event logs for building event classifiers; by examining the classification results, domain experts can further enhance the classification models. To the best of our knowledge, this paper is the first to study weakly supervised event classification in power system. The findings in the paper can shed the light on using imperfect real-world PMU data with scarce labels for event classification.

Using the two-year real-world PMU data from the Western Interconnection of the continental U.S. transmission grid, the performance of the proposed weakly supervised event classification approach is evaluated. The experimental results show that when only 5% of data are labeled, the average accuracy of the estimated labels using our approach can be around 70.9% and the average accuracy of the corresponding event classifier can achieve about 78.4%. This shows a promising way of using weakly supervised learning for developing robust event classification with extremely insufficient labeled data.

The rest of the paper is organized as follows: Section II provides the details of the proposed weakly supervised learning framework. Section III evaluates the performance of event classifiers trained under the proposed framework. Section IV concludes the paper.

II. Weakly Supervised Event Classification

A. Weakly Supervised Learning Framework

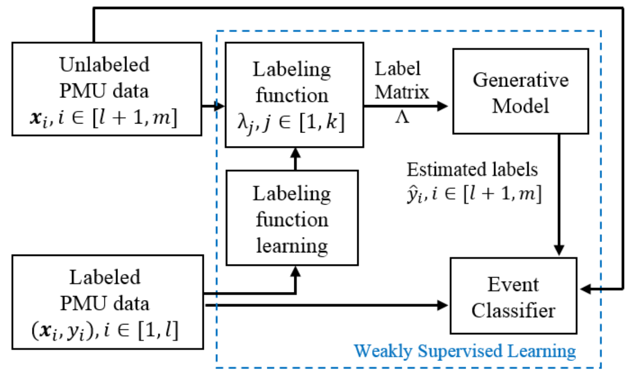

Fig. 1: The weakly supervised learning framework.

A weakly supervised learning framework (see Fig. 1) is proposed to estimate the labels of large amounts of unlabeled PMU data for training event classifiers. As the raw PMU data are often of low quality (e.g., bad data, dropouts, and timestamp errors), the data quality issues are initially fixed before using the proposed framework to estimate the labels, where the data preprocessing proposed in our recent work [1] is carried out and construct event features from the PMU data. Then, the labels of unlabeled PMU data are estimated in two main steps: labeling function learning and generative model based label estimation.

- **Labeling function learning.** Labeling functions (LFs) can be treated as event classifiers. Given the input PMU data, a LF outputs a label (i.e., event type), e.g., line outage, transformer outage, or frequency event. In the proposed framework, multiple LFs are learnt to characterize the features of different event types. Due to the limited labeled PMU data, the learnt LFs can be noisy and biased. Therefore, directly using the estimates from the LFs cannot generate good event classifiers.
- **Generative model.** To enhance the accuracy of the estimated labels from the LFs, a generative model is developed. As LFs are learnt using the same dataset, the estimated labels from the LFs are correlated. The goal of the generative model is to characterize the dependency structure of the LFs, based on which the estimated labels are better combined from the LFs to generate better estimates without knowing the ground truth.

Using the estimated labels from the generative model together with the limited labeled data, event classifiers are trained, where different off-the-shelf machine learning models (e.g., random forest) can be used. In the following, the details of each main step in the proposed framework are discussed.

B. Labeling Function Learning

From our previous work [1], it is observed that the patterns of the PMU measurement signals under different event types are distinct, as illustrated in Fig. 2. For example, frequency

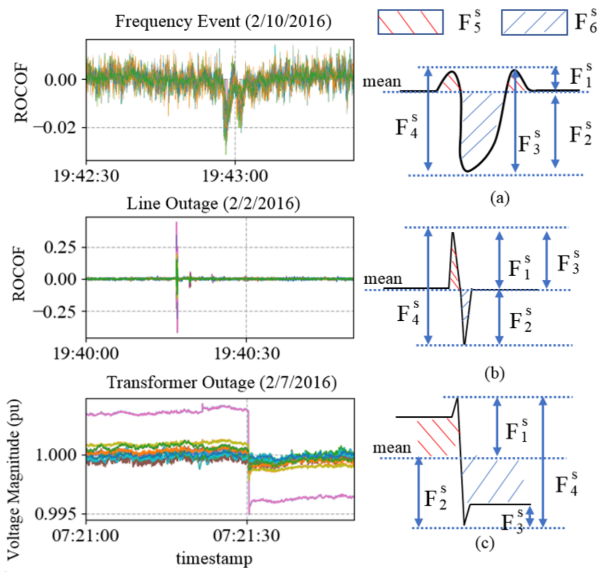

Fig. 2: Illustration of constructed features for different event types.

events have a deep and wide V-shape; line outages have a narrow spark; and transformer outages are similar to a step function. Based on these event characteristics, multiple event features F_i^s (i.e., Amplitude/Area above or below the average, Ramp up, Ramp down rate) are constructed [1] for each measurement signal s (including the voltage magnitude and the current magnitude of positive sequence and ROCOF) to train event classifiers. The event types are jointly determined by these features, not the single signal type. Because these features can better capture the patterns of different events, the event classifiers trained using these features are more robust and accurate than event classifiers trained using the PMU measurements directly [1]. Therefore, multiple LFs are learnt based on these features of different types of events.

Let $\mathcal{C} = \{(\boldsymbol{x}_1, y_1), ..., (\boldsymbol{x}_l, y_l), \boldsymbol{x}_{l+1}, \boldsymbol{x}_{l+2}, ..., \boldsymbol{x}_m\}$ denote the training dataset with m training samples, where $\boldsymbol{x}_i$ denotes the set of constructed features F_i^s of the training sample (event) based on our previous work [1]. In the training dataset $\mathcal{C}$, only l samples are labeled, where y_i denotes the event label for $i = 1, ..., l$. In practice, the number of labeled data is much smaller than the total number of data, i.e., $l \ll m$.

Using these l labeled samples, multiple LFs are trained using different machine learning models, e.g., Random Forest (RF), Logistic Regression (LR), K-Nearest Neighbor (KNN), and Gradient Boosting Decision Tree (GBDT). Specifically, event features constructed from different measurement signals (including voltage magnitude, current magnitude, and rate of change of frequency (ROCOF)) and their combinations are used to train different machine learning models. Table I shows the possible LFs to build using different features and models, where each LF is denoted as λ_j.

Suppose k LFs (i.e., $\{\lambda_j, j = 1, 2, ..., k\}$) are learnt using the l labeled samples. The output of each LF, i.e., $\lambda_j(\boldsymbol{x}_i)$ for $i = l+1, ..., m$, is the estimate of the event label of the unlabeled data. Let Λ be the label matrix generated by the k LFs for the unlabeled data, where $\Lambda_{i,j} = \lambda_j(\boldsymbol{x}_i)$ for $i \in [l+1, m]$ and $j \in [1, k]$. For a given unlabeled data $\boldsymbol{x}_i$, the outputs from k LFs (i.e., $\Lambda_{i,j}$ for $j \in [1, k]$) can be treated as noisy votes. From these noisy votes, it aims to estimate the "true" label without knowing the ground truth.

TABLE I
LFs BUILT BY DIFFERENT FEATURES AND CLASSIFIERS

Feature \ Classifier	RF	LR	KNN	GBDT
Voltage magnitude	λ_1	λ_2	λ_3	λ_4
Current magnitude	λ_5	λ_6	λ_7	λ_8
Rocof	λ_9	λ_{10}	λ_{11}	λ_{12}

As LFs are learnt using the same dataset with scarce labels, the estimated labels (i.e., $\Lambda_{i,j}$) are often correlated and the estimates can be noisy and biased. Simply using the majority vote to combine $\Lambda_{i,j}$ for $j \in [1, k]$ cannot obtain good estimate, as the majority vote assumes that the LFs are independent. To enhance the accuracy of the estimated labels, it is of paramount importance to characterize the correlations among the LFs, based on which $\Lambda_{i,j}$ for $j \in [1, k]$ can be better combined.

C. Generative Model

To better estimate the true labels Y using Λ, a generative model is leveraged, where the output of each LF, i.e., $\Lambda_{i,j}$, is modeled as a random variable. Using the observation matrix $\bar{\Lambda}$ (i.e., the outputs of Λ using the unlabeled data), the joint distribution $\text{Prob}(\Lambda, Y)$ of this generative model is estimated. $\text{Prob}(\Lambda, Y)$ will be used to estimate the accuracy of each LF, based on how well Λ are combined to estimate Y. Specifically, the joint distribution can be specified as:

$$\text{Prob}(\Lambda, Y) \propto \exp\left(\sum_{i=l+1}^{m} \sum_{d \in \mathcal{D}} \theta_d \phi_d(\Lambda_i, y_i)\right), \quad (1)$$

where $\Lambda_i = (\Lambda_{i,1}, ...\Lambda_{i,k})$ denotes the vector of the estimated labels from k LFs for the training sample i. $\mathcal{D}$ is a set of tuples that describe the dependencies between LFs, denoted by $\phi_d(\Lambda_i, y_i)$, and θ_d denotes how correlated the LFs are. For example, if d refers to the dependency between two LFs j_1 and j_2, then $\phi_d(\Lambda_i, y_i) = \phi_{j_1,j_2}(\Lambda_i, y_i) = \mathbb{1}\{\Lambda_{i,j_1}, \Lambda_{i,j_2}\}$, where the indication function $\mathbb{1}\{\Lambda_{i,j_1}, \Lambda_{i,j_2}\}$ represents whether LFs j_1 and j_2 depend on each other.

As Y is unknown for the unlabeled data, therefore the distribution $\text{Prob}(\Lambda)$ will be estimated. In [17], it is proved that using an unlabeled input dataset of size greater than a threshold, it is sufficient to recover the exact dependency structure. As the number of possible dependencies increases at least quadratically with the number of LFs, the approach in [17] is leveraged to efficiently learn the dependencies by changing the optimization objective to the log marginal pseudolikelihood of output of a single LF λ_j conditioned on the outputs of the other LFs $\lambda_{\backslash j}$ using L_1 regularization, i.e.,

$$\begin{aligned} &\underset{\theta=\{\theta_d, \forall d \in \mathcal{D}\}}{\arg\min} \; (-\log \; \text{Prob}(\bar{\Lambda}_j | \bar{\Lambda}_{\backslash j}) + \epsilon ||\theta||_1) \\ &= \arg\min_{\theta} (-\sum_{i=l+1}^{m} \log \sum_{y_i} \; \text{Prob}(\bar{\Lambda}_{i,j}, y_i | \bar{\Lambda}_{i\backslash j}) + \epsilon ||\theta||_1), \end{aligned} \quad (2)$$

where $\epsilon > 0$ is a hyperparameter. This problem can be solved efficiently using the stochastic gradient descent method [17].

After solving (2), the dependencies based on θ that have a sufficiently large magnitude to estimate $\text{Prob}(\Lambda, Y)$ are selected. For each unlabeled sample i, the estimated label $\hat{y}_i$ can be obtained by $\hat{y}_i = \arg\max_{y_i} \text{Prob}(y_i|\bar{\Lambda}_i)$ for $i = l+1, ..., m$. In this paper, a python package Snorkel [15] is applied to train this model.

D. Weakly Supervised Event Classification

Using the estimated labels based on the generative model, the large amount of unlabeled PMU data is leveraged to train event classifiers, where different machine learning models can be used. In this study, we compare the performance of different machine learning models using the training data with the estimated labels, i.e., $\{(\boldsymbol{x}_1, y_1), ..., (\boldsymbol{x}_l, y_l), (\boldsymbol{x}_{l+1}, \hat{y}_{l+1}), ..., (\boldsymbol{x}_m, \hat{y}_m)\}$. Based on the experiments (see Section III-C), it is observed that the Random Forest model performs best.

III. Case study of real-world PMU data

A. Experimental Setup

1) Data: This paper uses real-world PMU data from the Western Interconnection of continental U.S. transmission grid. The measurement data is collected from 23 PMU streams over a two-year period (2016–2017). The sampling rate of PMUs is 60 samples per second. The measurements used in the experiments are voltage magnitude of positive sequence, current magnitude of positive sequence, frequency, and rate of change of frequency (ROCOF). Besides, event logs for the period of two years are provided, where start timestamp, end timestamp, event type, event cause, and event description for each event are provided.

In the experiments, 3,877 events are used in the event logs, including 2,507 line events, 921 transformer events, and 449 frequency events. For these events, these events are randomly split in each month into 80% for training (i.e., 3098 events) and 20% for testing (i.e., 779 events), in order to ensure the events across the entire year are fairly distributed between the training set and the testing set. For the training data, 5% data are preserved as labeled data (i.e., 154 events) and remove the labels for the remaining training data, in order to evaluate the performance of the proposed weakly supervised event classification. The event classification performance have been tested over 20 runs, and in each run, the data are randomly selected.

2) Evaluation Metrics: Four metrics are used to evaluate the classification performance, i.e., accuracy (ACC), precision (PRE), recall (REC), and F1 score, which are defined as follows:

$$\begin{aligned}
\text{ACC} &= (\text{TP+TN})/(\text{TP+TN+FP+FN}), \\
\text{PRE} &= \text{TP}/(\text{TP + FP}), \\
\text{REC} &= \text{TP}/(\text{TP + FN}), \\
\text{F1} &= 2 \times (\text{PRE} \times \text{REC})/(\text{PRE + REC}),
\end{aligned}$$

where TP (i.e., True Positive) and TN (i.e., True Negative) denote the number of positive and negative instances that are correctly classified, respectively. FP (i.e., False Positive) and FN (i.e., False Negative) denote the number of misclassified negative and positive instances, respectively.

B. Performance of Weakly Supervised Event Classification

TABLE II
PERFORMANCE OF WEAKLY SUPERVISED EVENT CLASSIFICATION UNDER DIFFERENT MODELS USING THE TESTING DATA.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
RF	78.4±1.0	81.2±3.0	63.8±2.8	63.9±1.3
GBDT	77.5±3.3	77.2 ±7.5	62.6 ±10.2	64.8 ±8.5
SVM	71.1±0.1	35.7±21.4	33.63±0.9	28.3±2.0
LR	71.6±3.4	58.9±26.07	38.4±4.7	36.0±6.8
KNN	74.1±2.0	69.2±6.0	54.7±8.4	56.0±7.9
DT	74.9±3.3	70.6±7.3	62.1±8.4	64.3±6.8

Table II compares the average performance of different machine learning models over 20 runs, including Random Forest (RF), Gradient Boosting Decision Tree (GBDT), Support Vector Machine (SVM), Logistic Regression (LR), K-Nearest Neighbor (KNN), and Decision Tree (DT). Each model is trained using the training data with the estimated labels obtained using the weakly supervised learning and then tested using the testing data. It is observed that RF outperforms the other models in terms of accuracy, precision, and recall, and F1 score of RF is close to GBDT but with less variation. Thus, RF is used to train the weakly supervised event classifier.

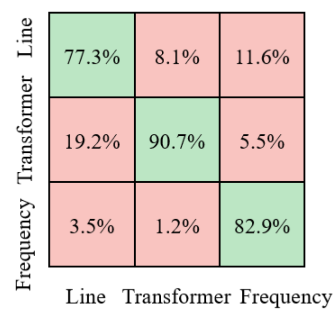

Fig. 3: Confusion matrix of the weakly supervised RF model for the testing data.

Fig. 3 illustrates the confusion matrix of the weakly supervised event classifier using RF, where the rows represent the estimated event type, the columns represent the true event type. The diagonal and off-diagonal cells in Fig. 3 represent the events that are correctly and incorrectly classified, respectively. From Fig. 3, it is observed that with only 5% labeled data, the proposed weakly supervised event classification can correctly classify the majority of the events (i.e., 78.4% on average). The misclassified events are due to the facts that 1) the patterns of line events and transformer events are similar and 2) the patterns in the ROCOF signal of some frequency events are buried by noise due to low signal-to-noise ratio (SNR). With scarce labeled data, it is challenging for the event classifier to correctly classify these events. It is believed that domain experts can further enhance the classification performance by examining the classification results, e.g., adding additional LFs using domain knowledge.

C. Weakly Supervised Learning vs. Semi-Supervised Learning

The performance of the proposed weakly supervised event classification with the semi-supervised learning based event classification approaches in [11] are compared. For the semi-supervised learning, the self-training under different models in [11] (i.e., LF, KNN, SVM, DT, Naive Bayes (NB), and the majority vote (Maj) based on these 5 models) are compared.

TABLE III
THE ACCURACY OF THE ESTIMATED LABELS UNDER DIFFERENT MODELS FOR THE 95% UNLABELED TRAINING DATA.

Model	Label Accuracy(%)
Our Method	70.9±0.6
Semi-DT	62.1±1.2
Semi-NB	64.7±0.3
Semi-SVM	64.7±0.1
Semi-LR	64.2±0.3
Semi-KNN	65.7±2.3
Semi-Maj	64.7±0.8

Table III compares the accuracy of the estimated labels using the proposed weakly supervised learning and the semi-supervised self-training under different models over 20 runs. It is observed that our method outperforms the semi-supervised models. Compared with the semi-supervised models, the proposed weakly supervised learning can easily incorporate the domain knowledge using different LFs, and thereby enhance the accuracy of the estimated labels.

TABLE IV
TESTING PERFORMANCE UNDER DIFFERENT MODELS.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
Fully supervised	94.3±0.1	95.1±0.3	87.9±0.2	91.0±0.2
Our Method	78.4±1.0	81.2±3.0	63.8±2.8	63.9±1.3
Semi-DT	69.9±1.4	74.3±1.2	56.0±2.5	48.9±2.5
Semi-NB	71.6±0.1	58.2±19.1	34.1±1.3	29.2±6.2
Semi-SVM	71.6±0.1	57.2±0.1	34.0±0.1	29.1±0.1
Semi-LR	71.3±1.5	54.4±25.7	35.2±9.9	31.3±13.8
Semi-KNN	73.3±0.5	79.5±5.9	42.8±10.8	43.6±13.7
Semi-Maj	71.6±0.3	58.9±31.7	34.1±1.4	29.2±2.7

Table IV compares the average performance of the weakly supervised event classification using RF and different semi-supervised approaches. The fully supervised learning case in our recent work [1] is also provided, where all the labels are provided in the training data. From Table IV, the proposed weakly supervised event classification outperforms the other semi-supervised models in all metrics by at least 5.1%, 1.7%, 7.8%, and 15% improvement in accuracy, precision, recall and F1 score, respectively. Compared to the fully supervised case, the weakly supervised event classification can obtain satisfactory results using only 5% labeled training data. Due to the scarce labeled data, the label estimation errors would degrade the performance of the event classification models. In the future, the use of other LFs developed using domain knowledge will be explored to enhance the classification performance.

IV. CONCLUSION

In this paper, a weakly supervised learning framework for training event classifiers using imperfect real-world PMU data with scarce labels is developed. One salient merit of the proposed framework is easy to incorporate the domain knowledge by adding labeling functions so that domain experts can further enhance the classification models. Another advantage is that less efforts are required from domain experts to maintain the event logs for building event classifiers. Numerical experiments using the real-world PMU data from the Western Interconnection of the U.S power transmission grid show that the event classifiers trained under the proposed framework when the training data has only 5% labeled data, a satisfactory classification accuracy of 78.4% is still achieved. In conclusion, the proposed framework offers a promising way to train event classifiers using scarce and low-quality labeled PMU data.

ACKNOWLEDGMENT

It is thankful for Pacific Northwest National Laboratory (PNNL) by offering the data in this study.

REFERENCES

[1] Y. Liu, L. Yang, A. Ghasemkhani, H. Livani, V. A. Centeno, P.-Y. Chen, and J. Zhang, "Robust event classification using imperfect real-world pmu data," *arXiv preprint arXiv:2110.10128*, 2021.

[2] I. Niazazari, Y. Liu, A. Ghasenikhani, S. Biswas, H. Livani, L. Yang, and V. A. Centeno, "Pmu-data-driven event classification in power transmission grids," in *2021 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pp. 1–5, IEEE, 2021.

[3] S. Liu, Y. Zhao, Z. Lin, Y. Liu, Y. Ding, L. Yang, and S. Yi, "Data-driven event detection of power systems based on unequal-interval reduction of pmu data and local outlier factor," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1630–1643, 2019.

[4] X. Zheng, B. Wang, D. Kalathil, and L. Xie, "Generative adversarial networks-based synthetic pmu data creation for improved event classification," *IEEE Open Access Journal of Power and Energy*, vol. 8, pp. 68–76, 2021.

[5] O. P. Dahal, S. M. Brahma, and H. Cao, "Comprehensive clustering of disturbance events recorded by phasor measurement units," *IEEE Transactions on Power Delivery*, vol. 29, no. 3, pp. 1390–1397, 2013.

[6] W. Li, M. Wang, and J. H. Chow, "Fast event identification through subspace characterization of pmu data in power systems," in *2017 IEEE Power & Energy Society General Meeting*, pp. 1–5, IEEE, 2017.

[7] Y. Chen, L. Xie, and P. R. Kumar, "Power system event classification via dimensionality reduction of synchrophasor data," in *2014 IEEE 8th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pp. 57–60, IEEE, 2014.

[8] H. Li, Y. Weng, E. Farantatos, and M. Patel, "An unsupervised learning framework for event detection, type identification and localization using pmus without any historical labels," in *2019 IEEE Power & Energy Society General Meeting (PESGM)*, pp. 1–5, IEEE, 2019.

[9] D. Senaratne, J. Kim, and E. Cotilla-Sanchez, "Spatio-temporal frequency domain analysis of pmu data for unsupervised event detection," in *2021 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pp. 1–5, IEEE, 2021.

[10] Q. Cui and Y. Weng, "Enhance high impedance fault detection and location accuracy via micro pmus," *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 797–809, 2019.

[11] H. Li, Y. Weng, E. Farantatos, and M. Patel, "A hybrid machine learning framework for enhancing pmu-based event identification with limited labels," in *2019 International Conference on Smart Grid Synchronized Measurements and Analytics (SGSMA)*, pp. 1–8, IEEE, 2019.

[12] Y. Zhou, R. Arghandeh, and C. J. Spanos, "Partial knowledge data-driven event detection for power distribution networks," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5152–5162, 2017.

[13] M. Farajzadeh-Zanjani, E. Hallaji, R. Razavi-Far, M. Saif, and M. Parvania, "Adversarial semi-supervised learning for diagnosing faults and attacks in power grids," *IEEE Transactions on Smart Grid*, 2021.

[14] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *National Science Review*, vol. 5, no. 1, pp. 44–53, 2018.

[15] A. Ratner, S. H. Bach, H. Ehrenberg, J. Fries, S. Wu, and C. Ré, "Snorkel: Rapid training data creation with weak supervision," in *Proceedings of the VLDB Endowment. International Conference on Very Large Data Bases*, vol. 11, p. 269, NIH Public Access, 2017.

[16] A. J. Ratner, C. M. De Sa, S. Wu, D. Selsam, and C. Ré, "Data programming: Creating large training sets, quickly," in *Advances in neural information processing systems*, pp. 3567–3575, 2016.

[17] S. H. Bach, B. He, A. Ratner, and C. Ré, "Learning the structure of generative models without labeled data," in *International Conference on Machine Learning*, pp. 273–282, PMLR, 2017.