Triple Generative Adversarial Networks Based Weather Classification Model for Short-Term Photovoltaic Power Forecasting

Yicheng Zhang, Ailing Zheng and Wen Yu

ABSTRACT

The challenge in the photovoltaic power (PV) forecasting is insufficiency of measuring dataset, especially for the extreme weather conditions. To alleviate this issue, Triple generative adversarial networks (Triple-GAN) based weather classification model is presented to predict short-term PV solar energy in a semi-supervised manner. The simulation results demonstrate that Triple-GAN can improve diversity of the data and capture the intrinsic features of the measuring weather data. Meanwhile, the accuracy of weather classification is improved and the weather classification model is robust for short-term PV forecasting.

KEYWORDS

Triple-GAN, Photovoltaic Power Forecasting, Weather Classification.

INTRODUCTION

Photovoltaic (PV) energy is one of the most significant and cleanest renewable resources since it is free cost and easily accessible. However, the characteristics of intermittence and fluctuation natures of PV solar energy have severe impact on the stable of the entire grid system. Thus accurate weather forecasting can mitigate the potential risk of high penetration rate. Recently weather classification model has been regarded as a significant role for enhancing the prediction accuracy[3, 4]. In addition, appropriate weather classification models fit for different weather conditions and can improve the prediction accuracy indeed, especially for the imbalance training data. However, few studies have focused on the weather classification. Since the extreme weather types are difficult to measure and obtain, the data of rare weather is scarce and will definitely cause the performance of PV forecasting deteriorate. Thus, imbalanced training data is a challenge for weather classification.

Recent advances in deep learning[1] have an explosive popularity in classification, which shows a potential to learn the discriminative representations of data by incorporating the feature extraction into the task learning process. Especially Generative
adversarial networks (GAN)[2] has attracted a lot of attention as it is capable of generating data without modelling the probability density function explicitly. It is intelligent for discriminator to incorporate unlabeled samples into training and impose higher order consistency by utilizing the adversarial loss. GAN has proven to be feasible in data augmentation, image-to-image translation, and semi-supervised learning (SSL). Our primary goal is to leverage triple-GAN to capture grid topology features of the daily solar irradiance data for weather classification in a semi-supervised manner. A novel Triple generative adversarial networks (Triple-GAN) has been developed. The generated data which capture the intrinsic features by resembling real historical weather data increase the diversity of generated samples. Balanced training data is able to describe the mapping relationship between different weather statuses without bias. Therefore, Triple-GAN is inherently suitable for improving the quality of generated weather data, balancing the amount of data in each weather type, and further enhancing the accuracy of weather classification model.

PROBLEM STATEMENT

Many generative models are developed in the literature. For instance, Douzas and Bacao (2018) presented a generative model titled “Random Oversampling (RO)”, which can generate new data efficiently but the training samples may lead to overfitting problems[3]. An approach of Synthetic Minority Over-sampling Technique (SMOTE)[4] is proposed to mitigate the overfitting problem, the datasets are balanced by simply up-sampling the rare class and down-sampling the majority class. However, the noisy is brought in SMOTE and lead to none-linear separation in the feature space since the boundary between minority and majority class is not often obvious.

Deep generative models (DGMs) are capable of capturing the underlying distributions of the data and generate new samples to resemble the realistic one. Over the last few years, significant progress has been made on generating realistic images by utilizing Generative Adversarial Nets (GANs)[2, 6, 7]. GAN is formulated as a two-player game, where the generator G takes a random noise z as input and try to produce a sample G(z) which looks like the real one as much as possible while the discriminator D identifies whether the sample comes from the generator G or the true data distribution p(x). Both G and D can be represented as deep neural networks and the objective function is to solve a minimax problem:

\[
\min_G \max_D U(D,G) = E_{x \sim p(x)}[\log(D(x))] + E_{z \sim p_z(z)}[\log(1-D(G(z)))]
\]  

Where \(p_z(z)\) is a simple distribution uniform or normal, and \(U(\cdot)\) represents the utilities. Given a generator G and the distribution \(p_\phi\) learnt by G, the optimal discriminator is \(D(x) = p(x)/(p_\phi(x) + p(x))\), and the global equilibrium of this two-player game can be achieved if and only if \(p_\phi(x) = p(x)\), which is desired for data generation. Furthermore, GANs have been proven to be effective in semi-supervised learning (SSL)[6], while the generative capability is retained. One more class
corresponding to the fake data generated by the generator is added to the categorical
discriminator for semi-supervised learning. There are two issues in existing GANs for
semi-supervised learning:(1) the generator and the discriminator (which also acts as
the classifier) may not be optimal simultaneously[5]; and(2) the generator cannot
make use of class label information and thus the semantics of the generated samples
can’t be controlled. To alleviate these issues, we exploit Triple-GAN, a flexible three-
player framework for weather classification in a semi-supervised manner.

MAIN PART

TRIPLE GENERATIVE ADVERSARIAL NETS

Inspired by[5], Triple-GAN is illustrated in Figure 1, Triple-GAN consists of three
components: a classifier, a generator, and a discriminator. (1) a classifier approximately
characterizes the conditional distribution 

\[ p_c(y|x) \approx p(y|x) \]

; (2) a generator
characterizes the conditional distribution in the opposite direction 

\[ p_g(x|y) \approx p(x|y) \];
and (3) a discriminator solely distinguishes whether the data-label pair \((x,y)\) comes
from the true distribution \(P(x,y)\). \(P(x)\) is the true data distribution of input irradiance
data and \(P(y)\) is denoted as the distribution of labels on annotated data. In the three-
player game, sample \(x\) is drawn from real data distribution \(P(x)\), classifier produces a
label \(y\) given \(x\) following the conditional distribution \(p_c(y|x)\). Therefore, the data-
label pair is a sample from the joint distribution \(p_c(x,y) = p(x)p_c(y|x)\). In the same
manner, a fake input-label pair can be generated from Generator by \(y \sim p(y)\) distribution
and then drawing \(x|y \sim p_g(x|y)\); hence form the joint
distribution \(p_g(x,y) = p(y)p_g(x|y)\). For \(p_g(x|y)\), we assume that \(x\) is transformed
from the latent variables \(z\) given the label \(y, x = G(y, z)\), \(z \sim p_z(z)\) where \(p_z(z)\) is a
simple distribution, here we use standard normal throughout the paper. Then, the fake
input-label pairs \((x,y)\) generated by both classifier and generator are fed into the
discriminator for judgement. Discriminator will identify the data-label pairs as positive
samples if it is from the true data distribution otherwise negative samples. We train
discriminator to maximize the probability of assigning the correct label to both real
examples and samples from generator \(G\) and classifier \(C\) respectively. To reach
equilibrium that the joint distributions learnt by the classifier and the class-conditional
generator both converge to the true data distribution, compatible objectives of
adversarial loss is defined as below:
Figure 1. An illustration of Triple-GAN.

\[
\min_{C,G,D} \max_{p(x,y)} U(C,G,D) = E_{(x,y)\sim p(x,y)} [\log D(x, y)] +
\alpha E_{(x,y)\sim p(x,y)} [\log(1 - D(x, y))] +
(1 - \alpha) E_{(x,y)\sim p(x,y)} [\log(1 - D(G(y, z), y))] 
\]

where \( C \), \( G \), and \( D \) is an individual network represented by vectorizing layers with parameters \( \theta_c \), \( \theta_g \), and \( \theta_d \) respectively; \( E(\cdot) \) denotes the expectation operator; \( D(x, y) \) denotes the probability that data-label pair came from the true data distribution; \( \alpha \in (0,1) \) is a parameter controlling the relative significance of generation and classification, and it is fixed as 0.5 in the experiments. The three-player game defined in Equation (1) reaches its equilibrium if and only if \( p(x,y) = \alpha p_c(x,y) + (1 - \alpha) p_g(x,y) \). The equilibrium implies that if one of classifier and generator goes towards the true data distribution, the other will also tend to true data distribution, which remedy the competing problem of the conventional two-player GAN. The equilibrium of \( U(C, G, D) \) is achieved if and only if \( p(x,y) = p_c(x,y) = p_g(x,y) \). Both classifier and generator will achieve the optimum simultaneously when they both converge to the true data distribution during the training process.

The discriminator loss is defined as:

\[
D_\text{loss} = \sum_{(x_d,y_d)} \log D(x_d, y_d) +
\alpha \sum_{(x_c,y_c)} \log(1 - D(x_c, y_c)) +
(1 - \alpha) \sum_{(x_g,y_g)} \log(1 - D(x_g, y_g))
\]

The classifier loss is defined as:
\[
C_\text{loss} = \alpha \sum_{(x_c, y_c)} p_c(y_c | x_c) \log \left( 1 - D(x_c, y_c) \right) + \mathcal{R}_c + \alpha_p \mathcal{R}_p
\]

(4)

The classifier loss is defined as:

\[
G_\text{loss} = (1 - \alpha) \sum_{(x_g, y_g)} \log \left( 1 - D(x_g, y_g) \right)
\]

(5)

The cross-entropy loss of real labelled data distribution is defined as \(\mathcal{R}_c = E_{(x,y) \sim \mathcal{P}(x,y)} [-\log p_c(y | x)]\), which is equivalent to model the KL-divergence between \(p_c(x, y)\) and \(P(x, y)\). As noted above, there is a lack of training dataset with respect to the extreme weather types. The generator can resemble such synthesis data with respect to the rare weather type for data augmentation purpose. The cross-entropy loss of synthesis data from class-conditional generator is defined as: \(\mathcal{R}_p = E_{p_g} [-\log p_c(y | x)]\), which optimizes classifier on the samples produced by generator in a supervised way. \(\alpha_p\) is the weight hyper-parameter fixed as 0.05 throughout the paper. Intuitively, a good generator can produce meaningful data-label pair beyond the scope of training set as auxiliary information for classifier, which will improve the weather forecasting performance, and vice versa, a good classifier will boost the performance of generator. As a result, both classifier and class-conditional generator can improve mutually. Moreover, the discriminator can utilize the label information of the unlabeled data through the classifier and then assist the generator to generate correct data-label pairs, which could alleviate mode collapse problem and is more likely to reach Nash equilibrium. Furthermore, minimizing \(\mathcal{R}_p\) with respect to classifier is equivalent to minimizing \(D_{KL}(p_g(x, y) \parallel p_c(x, y))\). Note that directly minimizing \(D_{KL}(p_g(x, y) \parallel p_c(x, y))\) is infeasible since the unknown likelihood ratio \(p_g(x, y) / p_c(x, y)\) cannot be computed directly.

**THE EVALUATION OF WEATHER CLASSIFICATION MODELS**

Weather classification models based on Triple-GAN are evaluated in the application domain of PV power forecasting. The data is download from the National Oceanic & Atmospheric Administration (NOAA) Earth System Research Laboratory website at Goodwin Creek in Mississippi during 2006 to 2008. The interval time is 3 min, and 1095 days are available. In order to meet the required standard of a short-period solar irradiance forecasting, the irradiance data is pre-processed to the time series within 15 min resolution. 96 data points are obtained for each day by taking the average of 5 points data in the span of every 3 min. During the night, the irradiance data is zero, thus the 18th to 78th data points for one day are chosen. The weather status is provided by Earth System Research Laboratory and we reclassified it into 9 classes corresponding to 9 weather types. The details of the 9 weather types are described in Table I. In the irradiance data of 1095 days, the data of the first 730 days is set as training samples and the irradiance data of the rest 365 days is set as testing samples. The number of samples and corresponding ration are shown in Table II.
| Weather types | Description | Weather types | Description |
|---------------|-------------|---------------|-------------|
| Morning       | Sunny       | Afternoon     | Sunny       |
| Class 1       | Sunny       | Class 6       | Cloudy      |
| Class 2       | Sunny       | Class 7       | Rainy       |
| Class 3       | Sunny       | Class 8       | Rainy       |
| Class 4       | Cloudy      | Class 9       | Rainy       |
| Class 5       | Cloudy      |               | Cloudy      |

### TABLE II. THE DISTRIBUTION OF NINE WEATHER TYPES IN TRAINING DATASET AND TESTING DATASET.

|          | Training dataset | Testing dataset |
|----------|------------------|-----------------|
|          | Samples | Ratio | Samples | Ratio |
| Class 1  | 225     | 30.82% | 110     | 30.14% |
| Class 2  | 91      | 12.47% | 45      | 12.33% |
| Class 3  | 38      | 5.21%  | 18      | 4.93%  |
| Class 4  | 39      | 5.34%  | 25      | 6.85%  |
| Class 5  | 75      | 10.27% | 35      | 9.59%  |
| Class 6  | 19      | 2.60%  | 10      | 2.74%  |
| Class 7  | 19      | 2.60%  | 11      | 3.01%  |
| Class 8  | 25      | 3.42%  | 12      | 3.29%  |
| Class 9  | 199     | 27.26% | 99      | 27.12% |

After data preprocessing and adding class labels corresponding to 9 weather types, the normalized solar irradiance data are fed into Triple-GAN. The input irradiance data is one-dimensional grid data at fixed time intervals. Meanwhile, irradiance is affected by some meteorological factors such as wind speed, wind direction, and cloud cover. In order to capture such complex factors, we modify the Triple-GAN with the following 4 significant improvements: (1) 2D transposed convolution is utilized in Generator to generate irradiance data. (2) DenseNet[10] is adopted in both classifier and discriminator to extract local features. (3) With respect to generator, the condition variable $y$ can either be concatenated with the random noise $z$ in the first layer or be added in the subsequent layers as additional channels. In our study, we adopt the later one. (4) As suggested by Radford et al.[7], we also add Batch Normalization (BN) to both the discriminator and the generator in Triple-GAN model to prevent generator from collapsing to a single point. However, adding BN to all layers causes model instabilities. Therefore, we also avoid using BN in the last output layer of generator and the first input layer of discriminator as they suggest.

Usually, when CNNs get deeper and deeper, the input gradient that passes through many layers can vanish at the end of the network. However, DenseNet utilizes dense connectivity to remedy this problem. It ensures maximum information flow through layers in the whole network by connecting all layers with each other directly using the same size of feature-map. Each layer requires information from all preceding layers as input, then it concatenates its own feature maps to all subsequent layers as illustrated in Figure 2. More details could be referred to[10].
The proposed Triple GAN architecture is trained on NOAA dataset from scratch in an end-to-end manner. Our model was implemented by TensorFlow and scikit-learn. It was run on NVIDIA GeForce GTX 1080 GPU. The initial learning rate is 0.01, and it will decrease to 0.001 at 75 epochs and 0.0001 at 110 epochs. A weight decay of 0.0001 has been applied for all the weights, and a Nesterov momentum[11] of 0.9 without dampening has been used in our model. The validation accuracy will be evaluated once for each training epoch. In addition, we set the batch size as 7, and the number of epochs as 150. The loss $R_p$ is not triggered until the number of epoches reach a threshold that the generator start to produce meaningful data. Here the threshold is set as 120, and $P$ is set as 0.05.

| Class | GAN/CNN2D | WGAN/GP/CNN2D | Triple-GAN/CNN2D | PA | UA | PA | UA | PA | UA |
|-------|-----------|---------------|-----------------|----|----|----|----|----|----|
| Class1| 1.000     | 0.961         | 1.000           | 0.867 | 0.990 |
| Class2| 0.550     | 0.941         | 0.800           | 0.815 | 0.946 |
| Class3| 0.462     | 0.909         | 0.769           | 0.545 | 0.917 |
| Class4| 0.211     | 0.727         | 0.421           | 0.571 | 0.857 |
| Class5| 0.556     | 0.615         | 0.593           | 0.441 | 0.692 |
| Class6| 0.200     | 0.429         | 0.600           | 0.500 | 0.667 |
| Class7| 0.200     | 0.667         | 0.400           | 0.250 | 0.500 |
| Class8| 0.714     | 0.750         | 0.429           | 0.714 | 1.000 |
| Class9| 0.846     | 0.753         | 0.938           | 0.753 | 0.838 |
| OA    | 0.753     | 0.842         | 0.858           |

Our proposed Triple-GAN is evaluated and compared with the other two representative GAN models conventional GAN and WGANGP. The other conditional factors are fixed to investigate the classification performance with respect to the 9 weather types separately. The comparison results are shown in Table III. It can be observed that Triple-GAN achieves better classification performance than the other two GAN models in terms of OA. It is worth noting that the sample size of Class 6 and Class 7 is extremely small due to the lack of data with rare weather types, thus the data distribution of the nine weather types is imbalanced. However, Triple-GAN can still achieve promising results in this application domain which indicates that Triple-GAN could generate synthesis samples to augment the data especially for the rare
weather type, thus the bias caused by imbalance data distribution is alleviated, and the classification results are boosted.

CONCLUSION AND FUTURE WORK

In this paper, we exploit the Triple-GAN for more accurate weather classification to further improve the performance of PV power forecasting. DenseNet architecture is utilized in classifier and discriminator of Triple-GAN to mitigate the gradient vanish problem. Extensive experimental evaluations on the NOAA database validate the effectiveness on significant classification performance improvement in the context of imbalanced data with respect to 9 weather types. The results also indicate that Triple-GAN can improve diversity of data to alleviate bias caused by imbalance data distribution.

ACKNOWLEDGMENT

This work was supported by the Overseas Study Program of Guangzhou Elite Project (GEP) in China.
Corresponding author: Wen Yu

REFERENCES

1. L. Y, B. Y, and H. G, “Deep learning,” Nature, no. 521, pp. 436–444, 2015.
2. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in Advances in Neural Information Processing Systems 27, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 2672–2680.
3. Douzas G, Bacao F., “Effective data generation for imbalanced learning using conditional generative adversarial networks,” Expert Systems with Applications, 2018, pp. 464-471.
4. Chawla NV, Bowyer KW, Hall LO., “SMOTE: synthetic minority over-sampling technique,” Journal of Artificial Intelligence Research, 2002, pp. 321–357.
5. C. Li, T. Xu, J. Zhu, and B. Zhang, “Triple generative adversarial nets,” in Advances in Neural Information Processing Systems 30, 2017, pp. 4088–4098.
6. T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, and X. Chen, “Improved techniques for training GANs,” in Advances in Neural Information Processing Systems 29, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, Eds. Curran Associates, Inc., 2016, pp. 2234–2242.
7. Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434, 2015.
8. Diederik P Kingma, Shakir Mohamed, Danilo Jimenez Rezende, and Max Welling. Semisupervised learning with deep generative models. In NIPS, 2014.
9. Augustus Odena. Semi-supervised learning with generative adversarial networks. arXiv preprint arXiv:1606.01583, 2016.
10. G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
11. I. Sutskever, J. Martens, G. Dahl, and G. Hinton, “On the importance of initialization and momentum in deep learning,” in Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28, ser. ICML’13. JMLR.org, 2013, pp. III–1139–III–1147.