A Method of Integrated Energy Metering Simulation Data Generation Algorithm Based on Variational Autoencoder WGAN

Penghe Zhang, Yang Xue, Runan Song*, Yining Yang, Cong Wang, Liu Yang
China Electric Power Research Institute Co., Ltd.
Haidian District, Beijing, 100192
*Corresponding author’s e-mail: songrunan@epri.sgcc.com.cn

Abstract—With the development of integrated energy systems, the penetration rate of all types of distributed power sources continues to increase. In a grid oriented to integrated energy metering, the metering business of different energy sources needs to fully consider the dynamic characteristics of different user loads. This paper aims at the generation of large-scale integrated energy metering data in large-scale integrated energy digital simulation technology, starting from supporting the construction of an integrated energy metering simulation system, combining various business applications and artificial intelligence technology under the background of the energy Internet. For the requirements of energy metering simulation, a method of integrated energy metering simulation data generation based on variational autoencoder WGAN is proposed, and the effect of the method proposed in this paper is verified through actual metering scene data.

1. INTRODUCTION
In recent years, with the continuous improvement of the level of clean and flexible energy production in our country, the demand for diversified and interactive energy consumption has become increasingly significant, and energy business development, technological innovation, and policy promotion are all facing new challenges. Integrated energy metering for multiple business application scenarios such as electric vehicle charging and replacement, customer-side energy storage, distributed energy access, and multi-meter integration collection is an important foundation for evaluating various energy utilization conditions and implementing energy conservation supervision and management in accordance with the law. Among them, the integrated energy metering system is inevitably a complex and diverse system, and energy companies have gradually increased their requirements for reliability, safety, and stability of the system. However, there is no systematic and large-scale research on simulation technology for energy metering equipment. The metering system simulations carried out by various energy metering industries are mostly physical simulations and field tests[1], large-scale verification has not yet been achieved, the simulation method is single, and the results of laboratory test, field test, and the effect of large-scale application are quite different, which cannot meet the verification requirements for the feasibility of various new services, new methods, and new technologies, so it seriously affects the fairness and justice of measurement, and has severely restricted the development and innovation of our country's energy measurement industry. Therefore, it is urgent to carry out the research of data generation technology in the digital simulation of large-scale integrated energy metering system to
provide simulation technology support for the construction of the system and the development of key
equipment.
In order to further promote the construction of a smart energy service system and meet the reliability,
safety, and stability requirements of various energy metering businesses, it is urgent to build a
comprehensive efficient, accurate and reliable energy metering digital and true hybrid simulation
system to achieve the true restoration of sundry scenes, and effectively support the exploration and
research of various new energy metering technologies. Aiming at the problem of generating large-scale
integrated energy metering data in large-scale integrated energy digital simulation technology, this
paper proposes a integrated energy metering simulation data generation method based on variational
autoencoder WGAN, starting from supporting the construction of an integrated energy metering
simulation system, combining with the needs of various business applications, artificial intelligence
technology and other key energy metering simulations under the background of the energy Internet, it
provides complete and sufficient data for integrated energy metering simulation in real time.

2. RELATED WORK
The generative confrontation network uses the generator to generate generated data with the same
distribution as the input data, and uses the discriminator to determine the gap between the generated
data and the real data, during the confrontation training process of the generator and the discriminator, it
gradually approaches the Nash equilibrium, so that the generated data is as similar as possible to the
original data[2]. The generative confrontation network improves the generation ability and robustness
of the model through the idea of game theory, and has a good data generation effect, but the generative
confrontation network does not have the function of specifically responsible for feature extraction, and
it is difficult to train[3].
Therefore, in order to solve the problem of training stability of the generative confrontation network,
this paper chooses WGAN to replace the GAN generative model. The variational autoencoding network
VAE can encode data into low-dimensional factors with original data characteristics[4]. Therefore, this
paper adds the coding network of the VAE encoder to the WGAN generator part for data feature
extraction, and proposes a method of data generation based on the variational autoencoder WGAN
(VAE-WGAN).

3. MATHEMATICAL MODEL

3.1 Objective Function
The simulation data generation model based on variational autoencoder WGAN is composed of three
parts: VAE encoder, WGAN generator and discriminator.
Assuming that the output of the hidden layer of the network is different in Gaussian way[5]:
\[
p(D_{l} (x) | z) = N(D_{l} (x) | D_{l} (\hat{z}), I)
\]
In the formula, Dis(x) is the l-th hidden layer of the network. Maximizing \(E_{q(z|x)}[\log p(x|z)]\) is
equivalent to maximum likelihood estimation, we can use \(E_{q(z|x)}[\log p(D_{l}(x)|z)]\).
Therefore, the objective function of the VAE encoder in the integrated energy simulation data
generation model is:
\[
\max_{\mathcal{E}_{\text{vae}}} \mathcal{E}_{\mathcal{E}_{\text{vae}}} \{ \log p(D_{l}(x)|z) - D_{\gamma}(q(z|x) | p(z)) \}
\]
(2)
The objective function of the WGAN generator in the integrated energy simulation data generation
model is:
\[
\max_{\mathcal{E}_{\text{gen}}} \{ \mathcal{E}_{\mathcal{E}_{\text{gen}}} [\log p(D_{l}(x)|z)] + E [\log (D(\hat{z}))] - E [\log (D(\hat{z}))] \}
\]
(3)
The objective function of the WGAN discriminator in the integrated energy simulation data
generation model is:
\[
\max_{E_{\mathcal{D}_{\mathcal{D}_{\mathcal{D}_{\mathcal{D}}}}} } \{ E [\log D(x)] + E [\log (1 - D(\hat{z}))] + E [\log (1 - D(\hat{z}))] \}
\]
(4)
3.2 Restrictions

In order to ensure that the generated measurement data can fit the actual value, it is necessary to meet the authenticity constraint and the similarity constraint.

The authenticity constraint is to ensure that the generated data can be close to the real scene, and the authenticity loss $L_r$ is defined as:

$$ L_r = W(G(z; \theta^{\phi}); D^{\phi}) $$

(5)

In the formula, $G(z; \theta)$ is the generated data of the generator; $W(\sim; \theta(D))$ represents the Wasserstein distance between the generated sample and the real sample.

$$ L_s = \|G(z; \theta^{\phi}) - I\| $$

(6)

In the formula, $I$ represents the real sample, and two-norm are used to measure the similarity of two matrices.

In summary, the final constraint for the generation of integrated energy metering data is:

$$ \min_{G(z; \theta)} L_r + L_s $$

(7)

Use Adam as the optimizer to optimize the latent variable $z$, so that the generated measured value is as close as possible to the real value. The final total sample is:

$$ \hat{i} = I + G(\hat{z}; \theta^{\phi}) $$

(8)

3.3 Network Structure

According to the parameters set for the network structure and the activation function in literature[6], the network structure of VAE-WGAN in this paper is shown in Fig.1, and the hyperparameter selection is determined by experimental tests. The detailed network parameters of the VAE-WGAN generator in this paper are shown in Table 1, and the detailed network parameters of the discriminator are shown in Table 2. In order to improve the training speed of the generator network, the ReLU activation function is used. The discriminator network is basically symmetrical with the generator network. In order to improve the recognition performance of the discriminator, LeakyReLU is selected as the activation function of the convolutional layer[7].

The calculation formula of the activation function is as follows[8]:

$$ \text{ReLU}(x) = \begin{cases} 
  x, & x > 0 \\
  0, & x \leq 0 
\end{cases} $$

(9)

$$ \text{LeakyReLU}(x) = \begin{cases} 
  x, & x > 0 \\
  0.2x, & x \leq 0 
\end{cases} $$

(10)

$$ \text{Sigmoid}(x) = 1 / (1 + e^{-x}) $$

(11)
4. CALCULATION EXAMPLE
The calculation example in this paper takes the user power data in the actual integrated energy metering scenario as an example to verify the effect of the VAE-WGAN data generation method proposed in this paper.

### TABLE 1. GENERATOR STRUCTURE PARAMETERS

| Name of network layer | Parameters                  | Numerical value |
|------------------------|-----------------------------|-----------------|
| Convolutional layer 1  | Size of convolution kernel  | 3               |
|                        | Number of filters           | 64              |
|                        | Step length                 | 1               |
|                        | Activation function         | ReLU            |
| Convolutional layer 2  | Size of convolution kernel  | 3               |
|                        | Number of filters           | 128             |
|                        | Step length                 | 1               |
|                        | Activation function         | ReLU            |
| Convolutional layer 3  | Size of convolution kernel  | 3               |
|                        | Number of filters           | 256             |
|                        | Step length                 | 1               |
|                        | Activation function         | ReLU            |
| Fully collected layer  | Number of neurons           | 1               |
|                        | Activation function         | sigmoid         |

### TABLE 2. DISCRIMINATOR STRUCTURE PARAMETERS

| Name of network layer | Parameters                  | Numerical value |
|------------------------|-----------------------------|-----------------|
| Convolutional layer 1  | Size of convolution kernel  | 3               |
|                        | Number of filters           | 64              |
|                        | Step length                 | 1               |
|                        | Activation function         | LeakyReLU       |
| Convolutional layer 2  | Size of convolution kernel  | 3               |
|                        | Number of filters           | 128             |
|                        | Step length                 | 1               |
|                        | Activation function         | LeakyReLU       |
| Convolutional layer 3  | Size of convolution kernel  | 3               |
|                        | Number of filters           | 256             |
|                        | Step length                 | 1               |
|                        | Activation function         | LeakyReLU       |
| Fully collected layer  | Number of neurons           | 1               |
|                        | Activation function         | sigmoid         |
4.1 Data generation process

The data generation process is shown in Fig. 2. The integrated energy metering simulation data generation method based on variational autoencoder WGAN includes the following process steps:

Step 1: Build an integrated energy measurement sample data set based on the existing integrated energy metering data;

Step 2: Use the VAE coding network of the variational autoencoder to extract the features of the integrated energy metering data in the sample library in step 1 to obtain the encoded data features;

Step 3: Input the data features encoded in step 2 into the WGAN network, and conduct confrontation training between the WGAN generator and the WGAN discriminator;

Step 4: Determine whether the data generated by the WGAN network in step 3 meets the authenticity constraint and the similarity constraint;

Step 5: Obtain the generated integrated energy metering simulation data set.

4.2 Training process

Process 1: A sample set is generated through the generation network, which is a fake label, and is used as the training set of the discriminator A together with the input real sample.

Process 2: Migrate the trained discriminator A parameters to the opposite discriminator B, use the generated sample set and the real samples in the training set to construct the mixed sample to train the discriminator B, update the network parameters, so that the discriminator A cannot distinguish the real samples from the generated samples.

Repeat the above two processes until the classifier A and classifier B cannot distinguish the authenticity of the sample, and the generated integrated energy metering simulation data is obtained.

The change process of the loss function of the generator and the discriminator during the training process of GAN and the VAE-WGAN network proposed in this paper is shown in Fig. 3 and Fig. 4.
Comparing the training process of GAN and VAE-WGAN proposed in this article, it can be seen that the generators of the two basically reach convergence after 150 iterations, and the discriminator basically converges after 300 iterations, but the loss functions of GAN generator and discriminator are oscillating, the convergence effect is not good.

In summary, through the change of the loss function curve of the generator and the discriminator, the training process of GAN is unstable, while VAE-WGAN has better stability.

4.3 Result of generated data
Fig. 5 shows the comparison between the generated data of the VAE-WGAN network after 30 rounds of training and the real samples. It can be seen from the figure that the gap between the generated data after 30 rounds of training and the real data is not too big, which is close to the distribution of the real samples, and can reflect the data characteristics of the real measurement scene.

5. CONCLUSION
The integrated energy metering simulation data generation method based on variational autoencoder WGAN proposed in this article can effectively generate simulation data in integrated energy metering scenarios, realize the reasonable generation of basic data in large-scale integrated energy metering simulation, and solve user’s data problems in the process of large-scale integrated energy metering online simulation. It can provide complete and sufficient data for large-scale integrated energy metering simulation in real time, and improve the rationality of data generation in the simulation process, so further improve the efficiency and effect of integrated energy simulation.
ACKNOWLEDGMENTS
This work is supported by General Science and Technology Project of State Grid Corporation of China-Research on Key Technologies of Digital and True Hybrid Simulation of Large-scale Integrated Energy Metering System.

REFERENCES
[1] CUI Yingbin, ZHANG Weixin, SHAN Zengli, et al. Systematic Test Method Based on Low-voltage Power Line Carrier Communication [J]. Proceedings of the CSU-EPSA, 2018, 30(01): 81-87.
[2] WANG Shouxian, CHEN Haiwen, PAN Zhixin, et al. A Reconstruction Method for Missing Data in Power System Measurement Using an Improved Generative Adversarial Network [J]. Proceedings of the CSEE, 2019, 39(01): 56-64+320.
[3] YANG Yulian, QI Linhai, WANG Hong, et al. Reconstruction of Missing Measurement Data in Distribution Network Based on Generative Confrontation and Dual Semantic Perception [J]. Automation of Electric Power System, 2020, 44(18): 46–56.
[4] ZHANG Wenqiang, CHEN Jing, YANG Wei, et al. Photovoltaic Uncertainty Modeling Based on VAE-CGAN [J]. Power System Technology, 2021, 45(04): 1273–1280.
[5] YANG C, WANG Z. An Ensemble Wasserstein Generative Adversarial Network Method for Road Extraction From High Resolution Remote Sensing Images in Rural Areas[J]. IEEE Access, 2020, 8: 174317–174324.
[6] XIAO Rui. Research on Data Generation Model Based on Generative Adversarial Network[D]. University of Electronic Science and Technology, 2019.
[7] Bao Ping, Liu Yunjie. Research on Fault Identification Based on Improved Deep Model in Combination of Generative Adversarial Networks Under Unbalanced Data Sets [J]. Journal of Electronic Measurement and Instrumentation, 2019, 33(03): 176-183.
[8] PATHAK D, KRAHENBUHL P, DONAHUE J, et al. Context Encoders: Feature Learning by Inpainting[C]//2016: 2536–2544.