An Improved CNN-Based Completion Method for Power Grid Middle Platform Data

Peng Wu*, Mingsheng Xu and Li Cheng
Jiangsu Electric Power Information Technology Co., LTD., LTD., Jiangsu Nanjing 210000, China

*Corresponding author email: cs_pengwu@163.com

Abstract. The transmission of power data would be likely to be interrupted or interfered, which results in the middle platform data missing. Missing data reconstruction plays a key role in power data processing, based on which the quality and utilization of power data have been enhanced. In traditional power data filling methods, only a single data distribution was considered, the correlation of power data in time and space was ignored. In this paper, an improved Convolutional Neural Network (CNN) method for filling power data was presented, and a CNN structure was designed. Through the unsupervised training of CNN, this method mines the correlation of data from the dimensions of time and space, and efficiently completes the missing data through the constraints of time continuity and space continuity. The completion results show that our method can fill the missing data efficiently, furthermore, experimental evaluations validate the performs of the proposed method.

Keywords: Middle platform data; Missing data reconstruction; CNN.

1. Introduction
The large amount of data collected by supervisory control and data acquisition (SCADA) system is of great significance to state evaluation, equipment assessment and system operation optimization. It is worth mentioning that we should take the quality of the power data collected by SCADA which is measured by the amount of missing data into consideration to obtain reliable and useful information. With the development of big data in recent years, the transmission, storage and analysis of massive big data in power grid have become the main research direction. However, the data of the middle platform contains many contributes of users, and it’s more likely to be missing in the real sample process. Therefore, it’s more urgent and significant to reconstruct the data of the middle platform. The research conclusions can’t reflect the characteristics and objective laws of power system operation correctly, unless the data used in research are based on real and reliable data collection. Therefore, we have to reconstruct the missing power data.

In previous studies, many scholars have used mathematical method such as mean filling, hot deck filling, regression filling and nearest distance filling to reconstruct missing data [1]. This kind of processing method only analyses from the perspective of data distribution, ignoring the timing characteristics and correlation in the power system, so the missing data reconstruction effect in middle platform isn’t very well. As an important branch of Deep Learning, we can use CNN to learn the consistency constraint between data in time and space, so as to better reconstruct the miss data in middle platform. In [2], the authors proposed a learning-based method LAI to estimate the missing data and they valid the algorithm by simulation. In [3,4], the authors presented a Deep BCD-net using identical encoding-decoding CNN to recover image. Both studies showed CNN’s ability in missing data
reconstruction. So we proposed a improve CNN-based method to reconstruct missing data in middle platform. Because the method presented in this paper are data-driven and don’t involve explicit modelling steps, the result can be highly accurate even if the scale of the missing data is huge. The main contributions of this paper are shown below:

- We describe the various forms of missing power data in middle platform and propose a double CNN to reconstruct the missing data.
- Many existing CNN designed for image processing is a two-dimensional and three-channel matrix, it can’t apply to the power data that is one dimension, we redesign the structure of CNN to fix it.
- Experiment evaluation on middle platform shows that the accuracy of the model is improved by our method.

2. Missing Forms of Power Data and Double CNN

2.1. Missing Forms of Power Data

Generally, power information collection system collects data once fifteen minutes, as we described above, missing data would result in massive data loss. For instance, in the system, −2 is applied to indicate that the value doesn’t exist. we can be able to conclude that the absence of data would affect the power data analysis. The quality of power big data can be improved through the reconstruction of missing data.

There are many forms of missing data, including continuous missing, discrete missing, a single attribute missing and multi-attribute missing, as shown in Table 1.

| Attribute | A1 | A2 | A3 | A4 | ... | An |
|-----------|----|----|----|----|-----|----|
| I1        | ?  | X12| ?  | ?  | ... | X1n|
| I2        | X21| ?  | X23| ?  | ... | ?  |
| I3        | X31| X32| ?  | ?  | ... | X3n|
| I4        | X41| ?  | X43| ?  | ... | X4n|
| ...       | ...| ...| ...| ...| ... | ...|
| In        | ?  | ?  | ?  | ?  | ... | ?  |

Where $X_{ij}$ represents the data, $I_i$ and $A_j$ represents the attribute variable and the ‘?’ represents the missing data. In [6], forms of missing data also includes horizontal data missing, vertical data missing and all attribute data missing. Single attribute data missing means that the data only missing on one attribute $A$ and the remaining attribute data is complete. Multiple attribute data missing is random. All the attribute data missing means that all data on a column or a row is missing. Aiming at reconstructing the missing power data in middle platform, this paper proposes an improved double CNN to fill the missing data.

2.2. Improved Double CNN

Double CNN presented in this paper consists of generator and discriminator, which are both CNN. As described earlier, the problem of missing data can analogous to the problem of filling missing image. [5] proposed the context and authenticity constraints that need to be satisfied in the reconstruction of missing image. Traditional CNN adopted JS divergence to measure the gap between real data and generated data, but our method adopted Wasserstein distance instead of JS divergence. In order to achieve ideal effect by minimizing Wasserstein distance as the aim of the training of double CNN. The successful application of CNN in image, natural language processing (NLP), speech recognition and other fields has proved that CNN can learn complex objective laws in unsupervised training. By applying CNN to power data reconstruction, CNN can learn the correlation between data from the history data, contributing to the improvement of the accuracy of the data reconstruction. Yet many exiting CNNs were designed for image processing, typical image input is a two-dimensional three-channel matrix, which is different from the input of power data. Thus, the existing structure of CNN
can’t apply to process power data. We must redefine the parameters of network, making it suitable for reconstruction of power data. The reconstruction of power data can be transformed into a generation problem with consistent context under the constraints of time dimension and space dimension. That means we can train a CNN generation model to generate power data, meanwhile, according to the data without missing, we need to train another CNN to discriminate whether the reconstruction of missing data form generator makes sense and select the most appropriate data for reconstructing. Structure of each individual CNN structure is shown as Figure 1.

![Figure 1. Structure of individual CNN.](image)

As mentioned above, we need to redesign the ordinary CNN to use it for power data reconstruction in middle platform. The network parameters of generator are shown in Table 2.

| Layers | Name                    | Parameter | Number   |
|--------|-------------------------|-----------|----------|
| 1      | Fully connected layer   | Neuron    | 128*6    |
|        |                         | Active function | ReLU   |
| 2      | ID upsampling           | factor    | 2        |
| 3      | ID convolutional layer  | Kernel    | 3        |
|        |                         | Num of wave filter | 128 |
|        |                         | Step      | 1        |
|        |                         | Active function | ReLU   |
| 4      | Normalization           | Momentum  | 0.8      |
| 5      | ID upsampling           | factor    | 2        |
| 6      | ID convolutional layer  | Kernel    | 3        |
|        |                         | Num of wave filter | 64   |
|        |                         | Step      | 1        |
|        |                         | Active function | ReLU   |
| 7      | Normalization           | Momentum  | 0.8      |
| 8      | ID convolutional layer  | Kernel    | 2        |
|        |                         | Num of wave filter | 4     |
|        |                         | Step      | 1        |
|        |                         | Active function | tanh  |

We have given the network structure of the generator, which is basically symmetric with the discriminator, but the difference is that the activation is replaced with Leaky ReLU to improve performance. Generally speaking, the differences between the improved CNN in this paper and the traditional CNN are as follows: we use two CNN networks for data reconstruction, one as a generator and the other as a discriminator and we redesigned the structure of CNN so that it can be used for the recovery of one-dimensional data. The parameters of CNN are shown in Table 2.
3. Methodology

We chose the data without missing in the history data as training data set. We suppose that there are $i$ types of data in the system, the corresponding value is $x_i$, we describe the complex correlation between data as $p_c(x)$. It is difficult to describe it through mathematical models. Suppose a set of noise vectors $y$, satisfying the joint Gaussian distribution $p_{y(y)}$. We can establish the mapping relationship between them thorough deep neural network. We can take the data which doesn’t need to be reconstructed and we know its distribution as input. The establishment process of mapping is realized by the training of double CNN.

As described above, we propose a double-CNN method, one of which is a generator and the other is a discriminator. output $G(y; \theta^{(G)})$ by generator and output $D(y; \theta^{(D)})$ by discriminator constitute the main component of the system, where $\theta^{(G)}$ and $\theta^{(D)}$ represent the weights of two kinds of network respectively. Input the noise $y$ into the generator, when we train the network. The distribution of the generated data $p_g(y)$ will fit the sample data $p_r(x)$ gradually. Discriminator would be trained at the same time when generator trained, and the data which input into discriminator includes the real sample data as well as the result from generator. Finally, discriminator will output the probability $p_r$, which means that input data is a real date in sample. The loss function of generator and discriminator are shown below:

$$L_G = -E_{y \sim p_r(y)}[D(G(y))]$$  \hspace{1cm} (1)

$$L_D = -E_{x \sim p_r(x)}[D(x)] + E_{y \sim p_y(y)}[D(G(y))]$$  \hspace{1cm} (2)

Where $E$ represents the expected distribution; $G(y)$ represents the output of the generator, $D(\cdot)$ represents the output of the discriminator network. Double CNN proposed in this paper is a zero-sum game problem essentially. The objective function of the process of game is shown below:

$$\min_G \max_D V(G,D) = E_{x \sim p_r(x)}[D(x)] - E_{y \sim p_y(y)}[D(G(y))]$$  \hspace{1cm} (3)

We can conclude from the above description and formulation that the generator attempts to generate data which are close to the true distribution pattern to mislead discriminator, making discriminator unable tell which data is from true sample data. After training, generator will get a distribution of the real sample data. Specially, the optimal goal in Eq.3 is measured by Wasserstein distance. The definition of the Wasserstein distance is shown below:

$$W(p_r,p_g) = \inf_{\gamma} \sum_{(x,y) \in \Gamma} E(\gamma)(x-y)^2$$  \hspace{1cm} (4)

Where $\prod(p_r, p_g)$ is the set of joint probability $\gamma$ with $p_r$ and $p_g$; $W(p_r,p_g)$ is the lower bound of expectation. When training discriminator network, a training data from history data is first constructed. Then calculate loss value and adopted Adam optimizer update parameters of network. Before each update, the update of discriminator network parameters will be executed to improve the speed of training. The reason why We can reconstruct power data by the CNN is that CNN can learn the time and space constraints in context, we assume that the loss of authenticity and similarity are $l_a$ and $l_s$, and the defines are shown as below:

$$L_a = D(G(y; \theta^{(G)}; \theta^{(D)}))$$  \hspace{1cm} (5)

$$L_s = \|G(y; \theta^{(G)}) \times M_s, I \times M_s\|_2$$  \hspace{1cm} (6)

Where $G(y; \theta)$ represents generated data of the generator, $D(\cdot)$ represents the output of the discriminator, i.e. the Wasserstein distance between the generated data and the real sample data. $L_a$ represents the output of the discriminator, i.e. the Wasserstein distance between the generated data and
the real sample data, \( \times \) in Eq.6 represents the multiplication of matrix elements and \( \mathbf{I} \) represents the measurement data with missing values. Therefore, the final optimization objective function is:

\[
\min_{y \sim p(y)} L_s + L_t
\]  

(7)

Taking Eq.7 as the optimization goal and using Adam optimizer to make generated data close to the missing data as far as possible. Finally, the reconstructed data consists of the available data in the original and the reconstructed data in the generated sample.

4. Experiment Evaluation

In this section, we use a power grid platform dataset collected by the state grid company for evaluation. In [7-10], they use the rate of data reconstruction, model accuracy generated by learning and iteration effect to measure the quality of their work respectively. However, they neglect to analyze the quality of data reconstruction methods under different data loss rates. We compare the method in this paper with the one that is well known, i.e. autoregressive (AR) and LSTM in different missing data rate. The evaluation index we used were RMSE and MAPE, which are computed as:

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{a} \sum_{j=1}^{b} (L_{i,j} - \hat{L}_{i,j})^2}
\]  

(8)

\[
MAPE = \frac{1}{n} \sum_{i=1}^{a} \sum_{j=1}^{b} \left| \frac{L_{i,j} - \hat{L}_{i,j}}{L_{i,j}} \right|
\]  

(9)

The experiments are run on a workstation with Intel i7-9700k CPU and 32G RAM. Before experiments, 5 to 25 percent of original data are randomly deleted to simulate the scene of missing. The final evaluation results are shown in Table 3 and Table 4.

**Table 3.** RMSE value of different method.

| Missing rate | 5%    | 10%   | 15%   | 20%   | 25%   |
|--------------|-------|-------|-------|-------|-------|
| AR           | 7.6538| 13.2592| 17.5362| 24.3603| 29.1243|
| LSTM         | 3.8653| 7.1684| 9.5641| 16.3294| 22.6438|
| proposed     | 2.4312| 2.9832| 8.1642| 12.6342| 15.3267|

Table 3 lists the RMSE values of all comparison methods. We can clearly see that the proposed method obtains the best RMSE value, which shows that the low-rankness of the data reconstructed by the method in this paper is strong.

**Table 4.** MAPE value of different method.

| Missing rate | 5%    | 10%   | 15%   | 20%   | 25%   |
|--------------|-------|-------|-------|-------|-------|
| AR           | 0.1613| 0.2431| 0.2795| 0.3821| 0.4165|
| LSTM         | 0.1296| 0.2114| 0.2394| 0.2764| 0.3311|
| proposed     | 0.0861| 0.1364| 0.1792| 0.2128| 0.2594|

The same conclusion can be drawn in Table 4. The propose method also achieved the lowest MAPE values. We can see that LSTM’s performance is also well, but LSTM’s training also needs a lot of data sets. The method proposed in this paper not only has higher data reconstruction rate, but also has better sequence prediction function.

5. Conclusion

In this paper, we presented an improved double CNN method to reconstruct the missing power data in middle platform. This method mines the correlation of data from two dimensions of time and space, and reconstructs the missing power data effectively through the constraints of time continuity and space continuity. We can conclude from the results in Table 3 and Table 4 that the method in this paper has a better effect. Especially in the case of high data loss rate, the proposed method has greater advantages.
Acknowledgements
This work was supported by the Science and Technology Project of State Grid Co., LTD (Research and Application of Service Design and Management Technology based on Data Middle Platform, 5700-202018181A-0-0-00).

References
[1] Miranda V, et al. Reconstructing missing data in state estimation with auto-encoders [J]. IEEE Transactions on Power Systems, 2012, 27(2): 604-611.
[2] Kim, M, et al. Learning-Based Adaptive Imputation Method with KNN Algorithm for Missing Power Data. Energies, 10(10), 1668.
[3] Q. Zhang, et al. Missing Data Reconstruction in Remote Sensing Image With a Unified Spatial-Temporal-Spectral Deep Convolutional Neural Network. IEEE Transactions on Geoscience and Remote Sensing, vol. 56, no. 8, pp. 4274-4288, 2018.
[4] Chun et al. Deep BCD-Net Using Identical Encoding-Decoding CNN Structures for Iterative Image Recovery. 2018 IVMSP, Zagorochoria, 2018, pp. 1-5.
[5] R. A. Yeh, et al. Semantic Image Inpainting with Deep Generative Models. 2017 IEEE CVPR, Honolulu, HI, 2017, pp. 6882-6890.
[6] Deng, W, et al. A missing power data filling method based on improved random forest algorithm. Chinese Journal of Electrical Engineering, 5(4), 2019, pp. 33–39.
[7] Crespo Turrado, et al. A New Missing Data Imputation Algorithm Applied to Electrical Data Loggers. Sensors, 15(12), 2015, pp. 31069–31082.
[8] Fan, G, et al. Lost data recovery for structural health monitoring based on convolutional neural networks. Structural Control and Health Monitoring, 26(10), 2019.
[9] Y. Bai, et al. Positive Active Power Missing Value Repairing with LSTM Prediction. 2020 12th IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), 2020, pp. 1-5.
[10] W. Shi et al. Improving Power Grid Monitoring Data Quality: An Efficient Machine Learning Framework for Missing Data Prediction, 2015 IEEE 17th International Conference on High Performance Computing and Communications, 2015, pp. 417-422.