2020 19th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)

Cascaded Denoising Convolutional Auto-Encoders for Automatic Recovery of Missing Time Series Data

Yuanyi Chen[1,2], Yubin Wang[2], Qiang Yang[1,2,3]
1.Zhejiang University NGICS Platform, Zhejiang University, 310027 Hangzhou, China
2.College of Electrical Engineering, Zhejiang University, 310027 Hangzhou, China
3.Jiangsu Key Construction Laboratory of IoT Application Technology, 214064 Wuxi, China
Email: qyang@zju.edu.cn

***Abstract*—This paper proposes a kind of supervised cascaded denoising convolutional auto-encoders (CDCAE), aiming to accurately recover the missing load data in electric power system. The one-dimensional load data are reshaped as two-dimensional image for data enhancement, which enables the convolutional neural network (CNN) to understand the semantics of load data. Numerical results in comparison with similar day filling (SDF) clearly validate the effectiveness of the proposed CDCAE in accuracy.**

***Keywords*-cascaded denoising convolutional auto-encoders; recovery of missing data; deep learning;**

I. Introduction

Big data technology has brought great vitality to the development of Energy Internet. It takes super large-scale data as the center, and requires comprehensive collection of massive data in the electric power system [1]. The data collected by State Grid have exceeded 60TB per day [2]. However, in the presence of massive data scale, due to terminal equipment failure, transmission channel disturbance and other interfering factors, the current sampling and communication systems inevitably suffer from the problem of data missing.

Generally, the missing data can be classified as continuous and discrete missing [3]. For the crucial measurement data, once there are long-term continuous missing segments, it can lead to serious safety incidents in power system. In the other hand, if there are random discrete missing points, it might result in reduction of signal to noise ratio (SNR), which degrades the accuracy of state estimation as well.

To address this issue, various forms of mathematical methods have been investigated and adopted, including SDF [4], polynomial interpolation [5], k nearest neighbor (KNN) [6], and other solutions (e.g., [7] [8]). But these algorithms are either too sensitive to boundary data, or unable to perceive data features from a large scale of time, which turns out poor robustness on long-term continuous missing situation.

This paper develops a supervised CDCAE model combined with a data reshaping strategy [9], to efficiently recover the missing load data through image inpainting technologies [10], as a sort of machine learning applications in smart grid [11]. The results based on the load data of Belgium grid (www.elia.be) confirm that this solution can obtain higher SNR than conventional SDF solutions, and performs well even at heavy missing rate.

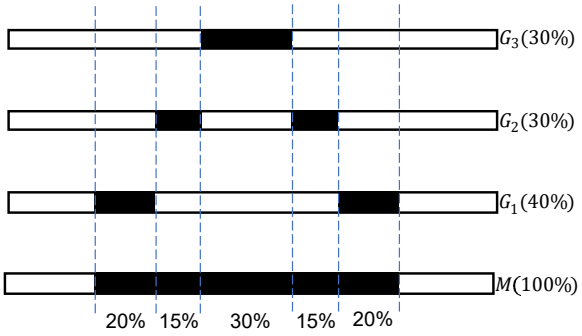

Figure 1. Mild (G_1), moderate (G_2) and severe (G_3) sub masks from M. The white regions represent normal data with m_i or $g_i^{(k)} = 0$,while black means missing data with m_i or $g_i^{(k)} = 1$.

II. Preprocessing

Missing data can be modeled via binary mask M as

$$X' = X - X \odot M \tag{1}$$

where $X = (x_1, x_2, \ldots, x_{n-1}, x_n)$ is the ground truth data, $M = (m_1, m_2, \ldots, m_{n-1}, m_n)$ is the binary mask in which m_i = 0 or 1, $\odot$ is the element-wise production operator and $X' = (x'_1, x'_2, \ldots, x'_{n-1}, x'_n)$ is the incomplete data which is regarded as input.

Define those normal data in X' as $\dot{X} = \{x'_i | x'_i \in X',$ and $m_i = 0\}$, and normalize X' as $\hat{X}' = (\hat{x}'_1, \hat{x}'_2, \ldots, \hat{x}'_{n-1}, \hat{x}'_n)$, and X as $\hat{X} = (\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_{n-1}, \hat{x}_n)$, where

$$\hat{x}'_i = (1 - m_i)\left(\frac{x'_i - \min(\dot{X})}{\max(\dot{X}) - \min(\dot{X})} 0.99 + 0.01\right) \tag{2}$$

$$\hat{x}_i = \frac{x_i - \min(\dot{X})}{\max(\dot{X}) - \min(\dot{X})} 0.99 + 0.01 \tag{3}$$

After that, 0 in $\hat{X}'$ will uniquely represent missing data while all the normal data range from 0.01 to 1. However, the normalized ground truth data in $\hat{X}$ might be out of range [0.01,1].

As Fig. 1 shows, the missing segments in M with $m_i = 1$, will be divided into three sub masks G_1, G_2 and G_3 as proportions of 40%, 30% and 30% and named as the

2473-3636/20/$31.00 ©2020 IEEE
DOI 10.1109/DCABES50732.2020.00080

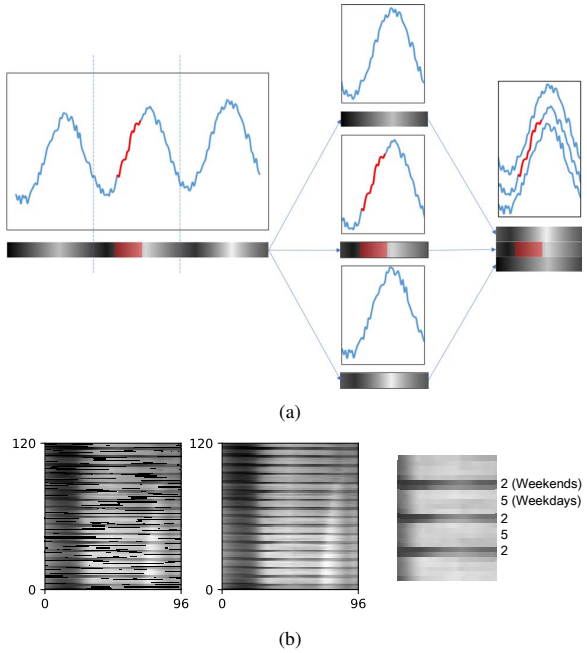

Figure 2. Visualization of data matrix by gray scale image. (a) diagram of a "generalized" image. (b) example of $A_{\hat{X}'}$ and $A_{\hat{X}}$.

mild, moderate and severe sub masks respectively, where $G_j = (g_1^{(j)}, g_2^{(j)}, \ldots, g_{n-1}^{(j)}, g_n^{(j)})$, and

$$G_1 + G_2 + G_3 = M \tag{4}$$

$$g_i^{(1)} + g_i^{(2)} + g_i^{(3)} = m_i \tag{5}$$

Assume the number of sampling points per day is m, and for dataset within k days set $n = km$. Then reshape $\hat{X}'$ ($\hat{X}$) as a k-by-m matrix $A_{\hat{X}'}$ ($A_{\hat{X}}$)

$$A_{\hat{X}'} = Reshaping(\hat{X}') = [a_{i,j}]_{k\times m} \tag{6}$$

where $a_{i,j} = \hat{x}'_{(i-1)m+j}$.

As Fig. 2(a) presents, since m is an approximate period of $\hat{X}'$, $A_{\hat{X}'}$ can be visualized through a gray scale image, which will be further considered as a "generalized" image. Fig. 2(b) left and middle demonstrate an example of gray scale images of $A_{\hat{X}'}$ and $A_{\hat{X}}$ with shape of 120-by-96 ($k = 120$, $m = 96$) from Belgium grid, where the regular light and dark gray stripes in the right corresponding to weekdays (high load) and weekends (low load) separately.

To provide more adjacent data for the left and right edges of $A_{\hat{X}'}$, two k-by-p redundant matrices $B_{\hat{X}'}$ and $C_{\hat{X}'}$ are designed as follow

$$B_{\hat{X}'} = [b_{i,j}]_{k\times p} \quad C_{\hat{X}'} = [c_{i,j}]_{k\times p} \tag{7}$$

where p is the padding depth and

$$b_{i,j} = \begin{cases} a_{i-1,m-p+j} & i > 1 \\ 0 & i = 1 \end{cases} \quad c_{i,j} = \begin{cases} a_{i+1,j} & i < k \\ 0 & i = k \end{cases} \tag{8}$$

Let $L = m + 2p$. Combine $A_{\hat{X}'}$ with $B_{\hat{X}'}$ and $C_{\hat{X}'}$, a k-by-L padding matrix $Z_{\hat{X}'}$ is generated as

$$\begin{aligned} Z_{\hat{X}'} &= Padding(A_{\hat{X}'}) = [B_{\hat{X}'}\ A_{\hat{X}'}\ C_{\hat{X}'}] \\ &= [z_{i,j}]_{k\times L} \end{aligned} \tag{9}$$

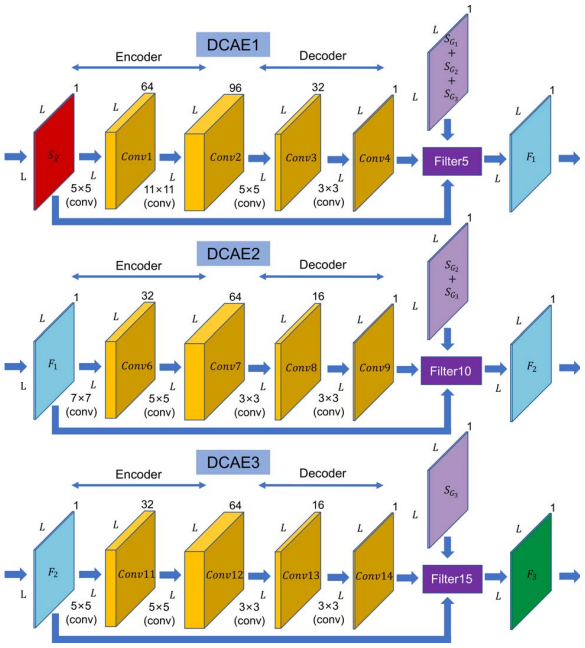

Figure 3. Structure of CDCAE.

To reduce the size, $Z_{\hat{X}'}$ will be separated into L-by-L slices (square submatrices) $S_{\hat{X}'}$, and the t-th slice is labeled as $S_{\hat{X}'}(t)$

$$S_{\hat{X}'} = Slice(Z_{\hat{X}'}) \tag{10}$$

$$S_{\hat{X}'}(t) = [s_{i,j}^{(t)}]_{L\times L} \tag{11}$$

where $s_{i,j}^{(t)} = z_{(t-1)m+i,j}$ and $t \le \lfloor \frac{k-2p}{m} \rfloor$ is met.

In addition,

$$Core(S_{\hat{X}'}(t)) = [y_{i,j}^{(t)}]_{m\times m} \tag{12}$$

is defined as the core area of $S_{\hat{X}'}(t)$, where $y_{i,j}^{(t)} = s_{i+p,j+p}^{(t)}$

Similarly, these operations will be conducted on $\hat{X}$ and G_k to obtain slices $S_{\hat{X}}$ and S_{G_k} respectively, which ensures the masks always hold the same shape as the data.

So far, the missing data can be regarded as noise or missing regions in "generalized" image, and moreover the original problem has been transformed into the issues of image inpainting and denoising.

III. Network Structure

As Fig. 3 illustrates, there are three denoising convolutional auto-encoders (DCAE1, DCAE2 and DCAE3) cascaded in CDCAE, corresponding to the three sub masks, where the missing data on S_{G_k} will be recovered k times.

Every DCAE has four convolutional layers in which the first two as encoder and the last two as decoder, followed by a filter layer. The padding model is 'same', the stride is one and the activation function is Relu for all those convolutional layers. The functions of filter layers are

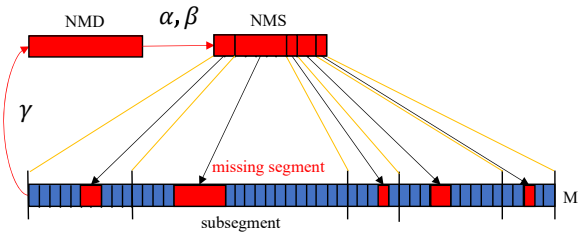

Figure 4. Generation of M.

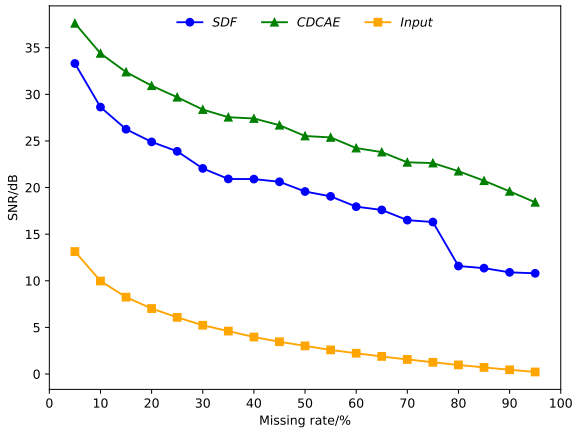

Figure 5. Comparison of SDF and CDCAE in SNR with respect to missing rate.

explained as follow.

$$F_1 = S_{\hat{X}'} + (Conv4 - S_{\hat{X}'}) \odot (S_{G_1} + S_{G_2} + S_{G_3}) \quad (13)$$

$$F_2 = F_1 + (Conv9 - F_1) \odot (S_{G_2} + S_{G_3}) \quad (14)$$

$$F_3 = F_2 + (Conv14 - F_2) \odot (S_{G_3}) \quad (15)$$

One of the most difference from the conventional DCAE [12] is that, here the encoder is upsampling and the decoder is subsampling, contrary to the general case. Besides, the decoder is made of convolutional layers instead of deconvolutional layers. Finally, the tensors (not only the output) always hold the same height and width as the input, i.e. L-by-L. These configurations enable the CDCAE to encode the data features without severe loss of semantics, and then recover the missing data as accurately as possible.

Specifically, S_{G_1} will be recovered once by DCAE1, S_{G_2} will be recovered twice by DCAE1 and DCAE2, and S_{G_3} three times by DCAE1, DCAE2 and DCAE3.

The loss function $\mathcal{L}$ is defined as the root mean square error (RMSE) on the core area of slices

$$\mathcal{L} = \sqrt{\frac{1}{\sum m_i} \|Core(F_3 - S_{\hat{X}})\|_2^2} \quad (16)$$

The final recovered data can be obtained simply by the reverse operation of the preprocessing in II.

IV. Numerical Results

Based on the load data of Belgium grid in every 15 minutes during 2014 to 2020, the proposed CDCAE will

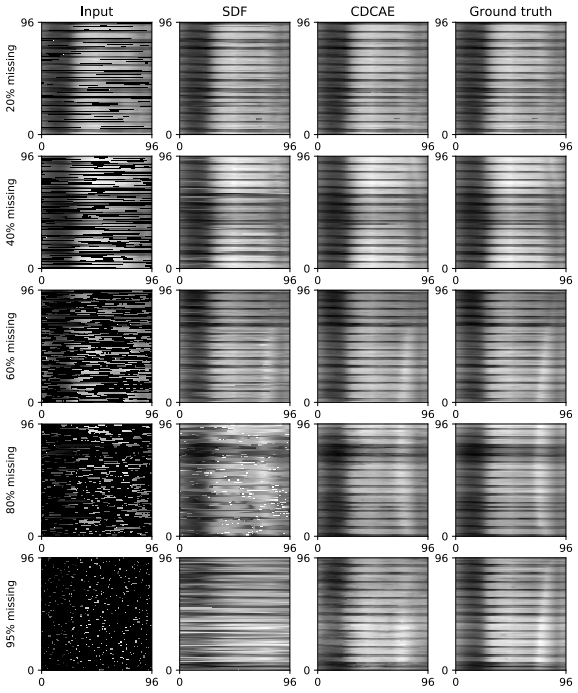

Figure 6. Input, output of SDF and CDCAE and ground truth in two-dimensional at missing rate of 20%, 40%, 80% and 95% (only the core).

be tested on different missing rate γ ranging from 5% to 95%, in comparison with SDF, where

$$\gamma = \frac{\sum m_i}{n} \quad (17)$$

Set $m = 96$ and $p = 9$, thus the size of slices is 114-by-114. The number of data slices is 540, in which 432 for training and 108 for testing. For both of the training and testing sets, the missing mask M is generated by stratified sampling, as Fig. 4 shows, in which

$$\text{NMD} = [n\gamma] \quad (18)$$

$$\text{NMS} = random([NMD\alpha], [NMD\beta]) \quad (19)$$

$$\text{ALMS} = \frac{2}{\alpha + \beta} \quad (20)$$

where NMD is the number of missing data, NMS is the number of missing segments, α and β are distribution coefficients ($\alpha \leq \beta$) which determine ALMS the average length of missing segments. Here α and β are fixed as $\frac{1}{120}$ and $\frac{9}{120}$, which means ALMS is 24 (a quarter day).

Consider the recovery error as noise, thus SNR is defined as

$$\text{SNR} = 20\lg\left(\frac{\|Core(S_{\hat{X}})\|_2}{\|Core(F_3 - S_{\hat{X}})\|_2}\right) \quad (21)$$

It is calculated with respect to missing rate 5% to 95%, for CDCAE, SDF as well as the input data before recovery.

Fig. 5 indicates the SNR results. The average SNR of CDCAE is 26.35dB while 19.63dB for SDF. It turns out that CDCAE outperforms SDF especially under heavy missing rate over 80%. Because then there hardly exists

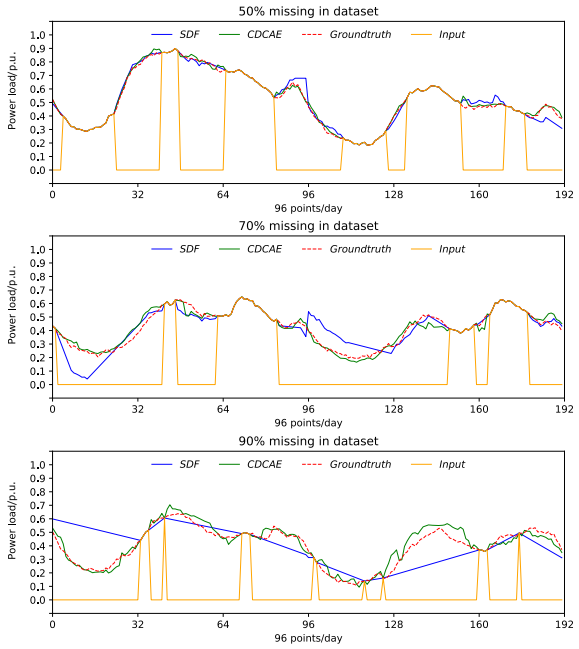

Figure 7. Input, output of CDCAE and SDF and ground truth in one-dimensional at missing rate of 50%, 70% and 90%.

any complete day without missing, which makes it nearly impossible for SDF to match similar days in dataset but only through linear interpolation and results in rapid drop at high missing rate. In the presence of extreme condition such as 95% missing rate, CDCAE is still able to keep a satisfied SNR more than 18dB from the input data with SNR less than 0.3dB, where it is only 10dB by SDF. In addition, CDCAE can be pretrained in advance and then used efficiently to process more than one hundred days simultaneously, but SDF is much slower because of the time complexity $O(n^2)$.

Furthermore, Fig. 6 displays the corresponding input, output of SDF and CDCAE, and ground truth in two-dimensional at different missing rate. Fig. 7 gives partial results reshaped back to one-dimensional. Obviously, CDCAE is much more powerful than SDF, and SDF will be easily degenerated as linear interpolation at high missing rate. These further evaluate the effectiveness of CDCAE.

V. Conclusion and Future Work

This paper presents a novel CDCAE model through reshaping the one-dimensional load data as two-dimensional image, to recover the missing data based on the principles of image inpainting, which guarantees a satisfied SNR even at extreme missing rate. Compared with the SDF solutions, CDCAE can comprehensively perceive and learn the features of load data in variant time scale, which ensures better performance. The numerical result confirms the validity on the proposed CDCAE model in accurately recovering the missing load data under different missing rate.

In future, the proposed solution needs to be further evaluated and validated through more real-time series datasets from various forms of application domains, e.g., renewable power generation, advanced metering and power quality condition monitoring.

Acknowledgment

This work is supported in part by the National Natural Science Foundation of China (Grant No. 51777183), the Fundamental Research Funds for the Central Universities (Zhejiang University NGICS Platform), and the Major Scientific Project of Zhejiang Lab (No. 2018FD0ZX01).

References

[1] S. Liu, D. Zhang and et al. A view on big data in Energy Internet[J]. Automation of Electric Power Systems, vol. 40, no. 8, pp. 14-21 and 56, 2016.

[2] T. Yang, F. Zhai, Y. Zhao, et al. Explanation and Prospect of Ubiquitous Electric Power Internet of Things[J]. Automation of Electric Power Systems, vol. 43, no. 13, pp. 9-20 and 53, 2019.

[3] J. Sun, Y. Jin, and M. Dai. Discussion on Testing the Mechanism of Missing Data[J]. Mathematics in Practice and Theory, vol. 43, no. 12, pp. 166-173, 2013.

[4] S. Zhao and C. Wang. Research and Appreciation of the Intelligent Recovery of Missing Data in Power System Measurement[J]. Science and Technology Innovation Herald,vol. 15, no. 18, pp. 96-98, 2018.

[5] S. Quan and Y. Cao. Interpretation Method Research Application[J]. Science & Technology Information, no. 36, pp. 413-414, 2007.

[6] P. Shi and L. Zhang. A Missing Data Complement Method Based on K-means Clustering Analysis[C]. IEEE Conference on Energy Internet and Energy System Integration(EI2), Beijing, 2017, pp.1-5.

[7] P. Gao ,M. Wang, S. G. Ghiocel and et al. Missing Data Recovery by Exploiting Low-Dimensionality in Power System Synchrophasor Measurements[J]. IEEE Transactions on Power Systems, vol. 31, no. 2, pp. 1006-1013, March 2016.

[8] V. Miranda, J. Krstulovic and et al. Reconstructing Missing Data in State Estimation With Autoencoders[J]. in IEEE Transactions on Power Systems, vol. 27, no. 2, pp. 604-611, May 2012.

[9] L. Wen, X. Li and et al. A New Convolutional Neural Network Based Data-Driven Fault Diagnosis Method[J]. IEEE Transactions on Industrial Electronics, vol. 65, no. 7, pp. 5990-5998, July 2018.

[10] D. Pathak, P. Krähenbühl and et al. Context Encoders: Feature Learning by Inpainting[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 2536-2544.

[11] M. Sohail Ibrahim, W. Dong, Q. Yang, Machine Learning Driven Smart Electric Power Systems: Current Trends and New Perspectives [J]. Applied Energy, vol. 272, August 2020, in press, doi: 10.1016/j.apenergy.2020.115237.

[12] P. Vincent, H. Larochelle, I. Lajoie and et al. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion[J]. Journal of Machine Learning Research, vol. 11, no.12, pp. 3371-3408, 2010.