

Data Augmentation Algorithm Based on Generative Antagonism Networks (GAN) Model for Optical Transmission Networks (OTN)

Liang CHEN^a, Kunpeng ZHENG^{b,1}, Yang LI^a, Xuelian YANG^a,
Han ZHANG^a, Yangyang LIANG^a, Junhua HUANG^a, Yuan
ZHANG^a, Pengfei XIE^a, Yongli ZHAO^b

^aInformation and communication branch of State Grid Corporation of China, China

^bBeijing University of Posts and Telecommunications, Beijing, China

Abstract. OTN (Optical Transmission Networks) is one of the mainstream infrastructures over the ground-transmission networks, with the characteristics of large bandwidth, low delay, and high reliability. To ensure a stable working of OTN, it is necessary to preform high-level accurate functions of data traffic analysis, alarm prediction, and fault location. However, there is a serious problem for the implementation of these functions, caused by the shortage of available data but a rather-large amount of dirty data existed in OTN. In the view of current pretreatment, the extracted amount of effective data is very less, not enough to support machine learning. To solve this problem, this paper proposes a data augmentation algorithm based on deep learning. Specifically, Data Augmentation for Optical Transmission Networks under Multi-condition constraint (MVOTNDA) is designed based on GAN Mode with the demonstration of variable-length data augmentation method. Experimental results show that MVOTNDA has better performances than the traditional data augmentation algorithms.

Key words. Optical transmission networks, GAN model, variable-length data, data augmentation, multi Condition constraint

1. Introduction

With the rapid development of optical networks, OTN play a very important role of data transmission in the ground, having the characteristics of large bandwidth, low

¹ Corresponding Author, Kunpeng ZHENG; E-mail: kunpengzheng@bupt.edu.cn.

delay, and high reliability [1]. OTN technologies become more and more mature with expanding of bigger network scale, especially including rapidly attractive difficulties of failure diagnosis. Driven by the era of big data, machine learning nowadays becomes a possible application solution for big data in communication fields, deriving various artificial intelligence algorithms. One typical application of machine learning is exploited for data learning in communication networks, with the deeply-mined rules that are hidden in the data. It has profound-guiding significance for the stable operation and safe production of the existing networks, in terms of data-based traffic analysis, alarm prediction, failure location, and construction of communication network equipment health portraits [2, 3].

However, the amount of effective data is small, and it is not enough to support big data-based algorithms such as machine learning. To solve this problem, it is generally introduced data augmentation algorithms to preprocess the data. While most of the new data augmentation methods are complex, mainly because most of them are based on machine learning models, given by the more common Logistic Regression classifier (LR) [4- 6], K-nearest Neighbors (KNN) and other algorithms [7, 8]. Then to the current hot development of Deep Neural Network (DNN) algorithm [9, 10]. From this, the idea of learning through game antagonism and GAN [11] is introduced to systematically extract features from scarce pre-processed data. On this basis, The AI model for data generation is introduced. Some researchers have introduced Recurrent Neural Networks (RNN) into GAN model, mainly because RNN is suitable for learning features from data sequences. Therefore, Continuous Recurrent Neural Networks with Adversarial Training (C-RNN-GAN) [12], Time-series GAN (Time GAN) [13] based on C-RNN-GAN supplemented with the data static characteristics, Doppel GANger (DG) [14] that associates metadata and feature data have appeared successively. Reference [15] proposed an adaptive traffic data enhancement algorithm is based on the generative confrontation network, enhanced the original traffic data from three different real scenes, and verified the rationality and dynamics of the proposed algorithm through the three statistical characteristics of mean value, variance and Hurst index. Literature [16] also proposed a self-optimized data enhancement algorithm based on original GAN for the enhancement of large optical transmission network alarm data, and verified the effectiveness of the enhanced data through alarm prediction scenarios.

This paper focuses on the data enhancement technology over self-optimization communication networks, and completes the research on the multi-condition variable length data augmentation method. Through the above research content, the “conditional, variable length, and high accuracy” augmentation of the scarce data of the optical transport network has been realized, which provides data support for the performance improvement of artificial intelligence model in communication network scenarios.

2. GAN Module

The GAN model has the core idea of generating antagonism, consisting of Discriminator Module (DM) and Generator Module (GM) two important structures. As shown in Figure 1, it is the basic structure of GAN model. The objectives of DM and

GM are opposite, thus presenting the dual antagonism of training process of GAN. The overall task of GAN for the data generation is to acquire the knowledge of data characteristics through the learning. Then, the synthetic data is used to identify the distribution law of original data, and pass for truth according to the learned knowledge. The objective function of GAN is described in Equation (1).

$$\min_G \max_D V(D, G) = E_{x \sim P_{data}(x)} [\log(D(x))] + E_{z \sim P_z(x)} [\log(1 - D(G(z)))] \quad (1)$$

$P_{data}(x)$ represents the distribution law of the real data set. $G(x)$ represents the process of GM processing input data, such as the target multidimensional data transformed by the random Gaussian white noise vector z into. $D(x)$ represents the probability that data x complies with the distribution law P_{data} of the real data set rather than P_g of the composite data set.

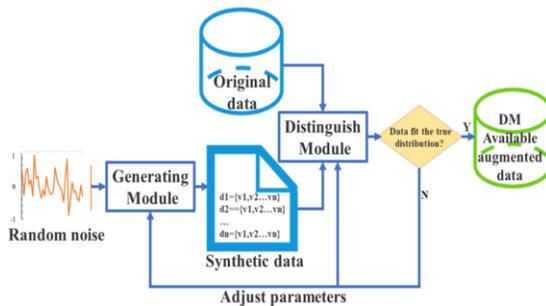


Figure 1. The basic structure of the GANs model.

3. VARIABLE-LENGTH DATA AUGMENTATION ALGORITHM FOR OTN UNDER MULTIPLE CONDITIONS (MVOTNDA)

This section describes a timing data generator, i.e., the DoppelGANger (DG) [17]. As shown in Figure 2, it is one architecture diagram of DoppelGANger model which is adapted and applied in OTN data augmentation.

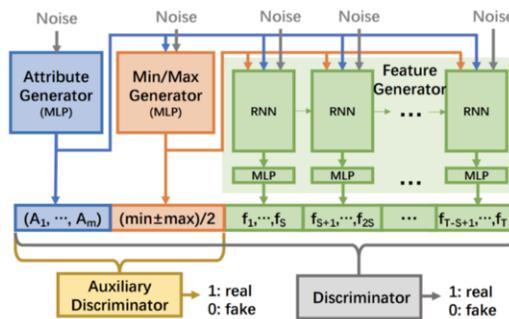


Figure 2. Doppel GANger model architecture diagram.

We conduct the DG model based on GANs and the Generator composed of RNN model, and the improved DG model effectively solves the limitations of timing data

Generator Module (GM). First, DG is an antagonistic generation network based on the GAN model. It uses Recurrent Neural Network (RNN) network model to improve the original GAN to synthesize more complex time series of data sets. For the application of DG, two innovations will be introduced to create the OTN data generator, i.e., a mature learning strategy and the mode collapse of GAN model, in DG to accelerate the convergence rate of GANs. Distinguishment Module (DM) adopts a well-designed noise mode processing strategy to improve the stability of GAN model.

As shown in Figure 3, one method is attribution separation generation method. The other method is a batch generation method, using to synthesize a small batch of long time series data sets. Then, a normalization method can be used to add normalization factor to GM to limit the range of data features. On the basis of DG, Convolutional Neural Networks (CNN) is introduced to analyze and recognize pictorial labels. It mainly aims at the primary optimization of topological data in OTN. MVOTNDA can augment the link data through the specified network links. The DG model is based on GAN's DG model and the generator composed of RNN model. Thus, the limitation of GM module of timing data can be solved effectively.

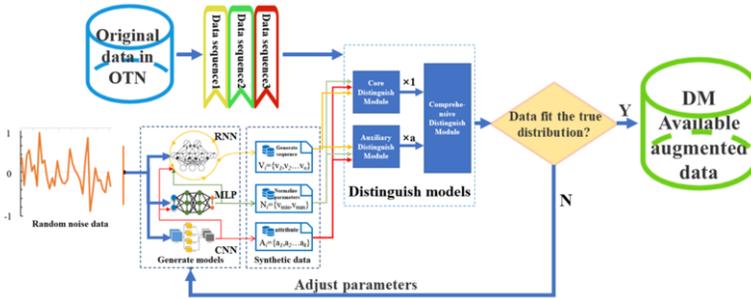


Figure 3. The block diagram of MVOTNDA algorithm based on DG model.

3.1 Objective optimization function of MVOTNDA

In this section, we designed the objective optimization function of MVOTNDA, Generation Module of MVOTNDA and Distinguishment Module of MVOTNDA. Reference [17] introduces Wasserstein distance to solve the model instability problem in the original GAN training process. When Wasserstein GAN (WGAN) model is trained, the adjustment of GM and DM training round ratio is omitted. The introduction of Wasserstein distance is mainly reflected in the following two modifications of the objective function,

(1) The logarithm function in the objective function is omitted;

(2) Each iteration update process of GM and DM, the model's relevant parameters were taken and deleted to keep them within a certain range.

To make the objective function conform to the problem, we convert the equation into the following form:

$$WL(G, D) = \min_G \max_D E_{s \sim P_{DS^k}} [D(s)] + E_{z \sim P_Z(z)} [-D(G(z))] \tag{2}$$

$G(z)$ is the new sample data generated by GM module. $D(s)$ is the value output by DM, represented the probability value that a generated sample passed by the GM to the DM will be identified as true.

P_{DS}^k represents the mathematical distribution of the original sample data set, P_Z represents the mathematical distribution of pseudo-data generated by the GM, $\min G$ and $\max D$ represent a process of simultaneous maximum and minimum optimization.

The detailed procedures of GAN model based on Wasserstein distance objective optimization function are shown in Table 1.

Table 1. Training process of GAN model based on Wasserstein distance objective optimization function.

Training process of GAN model based on Wasserstein distance objective optimization function
<p>Training parameter setting: Learning Rate, α ; Gradient Clipping, C ; Number of iterations of training, \mathbf{n} ; Number of iterations of the distinguish module, \mathbf{m} .</p> <p>Initialize i and j to 0</p> <p>While i < n do</p> <p> While j < m do</p> <p> Take any k $\{z^1, \dots, z^k\}$ input vectors from the white noise distribution $P_z(z)$;</p> <p> K data columns are randomly selected from the mathematical distribution P of the original data set; The DM was updated by Gradient Descent method:</p> <p> If the data is measured data:</p> $d_w := \nabla_{\theta_d} \frac{1}{k} \sum_l^k [-D(s^l) + D(G(z^l)) - \lambda(\ D_2(M(\hat{s}))\ _2 - 1)^2]$ <p> Else The data is label data</p> $d_w := \nabla_{\theta_d} \frac{1}{k} \sum_l^k [-M(D_2(s^l)) + D_2(M(G_2(z^l))) - \lambda(\ D_2(M(\hat{s}))\ _2 - 1)^2]$ <p> $w := w + \alpha \cdot d_w$</p> <p> $w := clip(w, -c, c)$</p> <p> j++;</p> <p> End While</p> <p> Take any k $\{z^1, \dots, z^k\}$ input vectors from the white noise distribution $P_z(z)$;</p> <p> The GM is updated by Gradient Descent method:</p> $g_w := \nabla_{\theta_g} \frac{1}{k} \sum_l^k [-D(G(z^l))]$ <p> $w := w + \alpha \cdot g_w$</p> <p> i++;</p> <p>End While</p>

As shown in Figure 2, it shows the importance to adjust the objective optimization function based on the Wasserstein distance. We introduce the Auxiliary Distinguishment Module (ADM) in the DM. The objective optimization function of MVOTNDA algorithm is designed as follows:

$$\min_G \max_{D_1, D_2} V(D, G) = WL(G, D_1) + \gamma WL(G, D_2) - \lambda \left(\|D_2(M(\hat{s}))\|_2 - 1 \right)^2 \tag{3}$$

Due to the unknown influence of ADM, we introduce a coefficient item γ to adjust the parameters during the experiment.

3.2 Generating Module of MVOTNDA

As shown in Figure 2, the GM of MVOTNDA algorithm is mainly used to synthesize the false data that follows the mathematical distribution of the original sample set. It is composed of three neural networks, namely CNN, MLP and RNN. Among them, RNN is responsible for the generation of pseudo-data, MLP is responsible for the generation of normalized parameters, and CNN is responsible for the generation of label data.

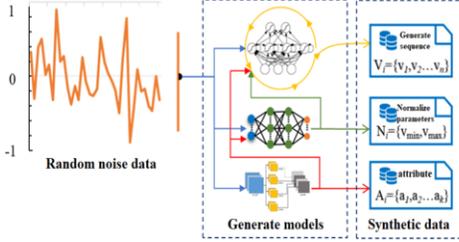


Figure 4. The GM structure.

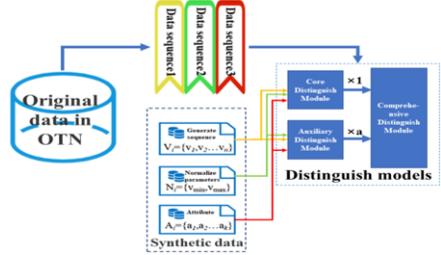


Figure 5. The DM structure.

CNN can identify the characteristics of each link in the topology and generate the corresponding data of each link, such as the flow data in the link. The output of CNN network acts on MLP normalized parameter generator to guide MLP to synthesize normalized parameters of data types identified by CNN. Finally, the output of CNN and MLP act on RNN together to guide RNN to generate sequence data under the corresponding label.

The GM structure of the MVOTNDA model is shown in Figure 4. By accepting the input of random white noise, the RNN model processes it to generate a sequence of side length:

$$\tilde{s}' = G(z) \tag{4}$$

Among them, z is a random white noise vector; \tilde{s}' represents a synthetic pseudo sequence whose length is variable. The objective optimization function of GM can be:

$$G_{loss} = E_{z \sim P_z(z)}[-D(G(z))] \tag{5}$$

$$G_{loss,2} = E_{z \sim P_z(z)}[-D_2(G_2(z))] \tag{6}$$

The loss value G_{loss} of the generated module needs to be minimized to maximize $D(G(z))$, so that the identification module cannot identify the authenticity of the data synthesized by the GM. the GM adopts autoregressive strategy. All output vectors adjusted by the fully connected neural network are merged into a time series in order to generate new and complete time series data.

3.3 Distinguishment Module of MVOTNDA

The DM is used to judge the authenticity of the input sequence data. Take a sequence from the original data sample set or the sequence synthesized by the generation module as input, and let the identification module determine the probability value of its output value between [0, 1]. The objective optimization function of the MVOTNDA algorithm's distinguishment module is shown at below:

$$D_{loss} = E_{s \sim P_{DS}(s)} [-D(s)] + E_{z \sim P_z(z)} [D(G(z))] \quad (7)$$

In the above equation, it is necessary to minimize the loss value D_{loss} of the identification model. This process mainly maximizes $D(s)$ and minimizes $D(G(z))$. In the other words, in the process of minimizing the objective optimization function, if the sequence in the original sample set is taken as the input, the value of $D(s)$ output by the identification module gradually approaches 1; if the input is the data synthesized by the GM, the output value $D(\tilde{s})$ tends to 0. After the training, the identification module can accurately determine the input sequence data. The specific structure of the identification module is shown in the Figure 5.

4. EXPERIMENTAL CONDITIONS AND RESULTS

The prediction is based on the augmentation and prediction of the traffic data of the British academic backbone network[18]. The data set used in the augmentation and prediction experiment of the traffic data of the British academic backbone network is from The UK Academic Network Backbone. The data traffic is mapped to the preset links in the topology. The experimental simulation topology is shown in Figure 6.

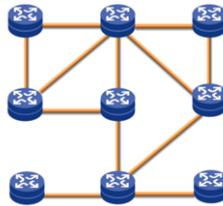


Figure 6. The simulation topology.

MVOTNDA is used to present the performances both of MD-SET and A1H. In addition, LSTM and the Original GAN model are used as the baseline. The Data conducted by the three models are shown in Table 2.

Table. 2 The comparison of three methods' performances.

Performance Index	MVOTNDA	LSTM	Original GANs
Similarity(SS)	0.976	0.826	0.865
Mutual Information(MI)	0.956	0.762	0.799
Accuracy rate	0.980	0.859	0.875

The accuracy of the data set is shown in Figure 7. It can be seen from the figure, most of dirty/redundant data is deleted after the preprocess. Only the performance sequence associated with alarm is left to generate a Data Set called the OD-set (Original Data Set). MVOTNDA augmented data set is MAD-set, with the LSTM-augmented data set is LSTM-set and the data generation of the original GANs model. Figure 8 shows the traffic prediction accuracy curve based on the three generated data. Followed by LSTM, the original GANs model is the worst one, while the accuracy of traffic prediction model trained by enhanced data of MVOTNDA is close to 97.5%.

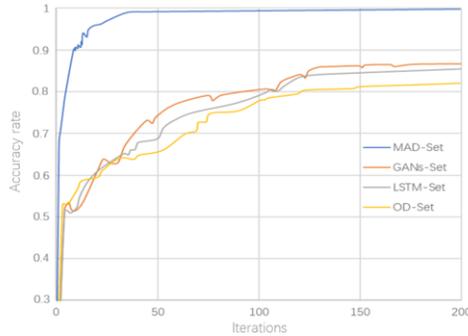


Figure 7. The accuracy of multi-alarm prediction model.

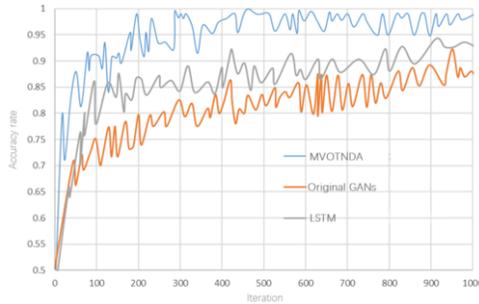


Figure 8. The accuracy of traffic prediction model.

5. Conclusion

In this paper, MVOTNDA is proposed under the background of OTN. Experiments show that augment traffic data in MVOTNDA is based on the specified link and the conditional generation capability. By introducing the label generation module and the auxiliary identification module, MVOTNDA can generate label data under the semi-supervised mechanism and act as the measurement data generation module MLP, which is also supervised by the core identification module, making the synthetic measurement data under the label conform to the mathematical distribution law of the original data set. MVOTNDA can quickly improve the accuracy to more than 90%, and the algorithm has good performance.

Acknowledgments

This work has been supported in part by National Natural Science Foundation of China (NSFC) (62021005, 61822105).

Reference

- [1] Takuya Ohara, Mitsuhiro Teshima, Shigeki Aisawa and Masahiko Jinno, "OTN Technology for Multi-flow Optical Transponder in Elastic 400G/1T Transmission Era"[J], OFC 2012, March 4, 2012- March 8, 2012
- [2] Zhao Y., Yan B., Liu D., et al. "SOON: self-optimizing optical networks with machine learning"[J]. Optics Express, 2018, 26(22):28713.
- [3] Zhang B., Zhao Y., Rahman S., et al. Alarm classification prediction based on cross-layer artificial intelligence interaction in self-optimized optical networks (SOON)[J]. Optical Fiber Technology, 2020, 57:102251.
- [4] Menard, Scott. "Applied logistic regression analysis.[M]. Sage Publications, Incorporated, 2001, Vol.106.
- [5] Breiman, Leo. Random forests[J]. Machine learning, 2001, 45(1): 5-32.
- [6] Jennie, Pearce, and, et al. Evaluating the predictive performance of habitat models developed using logistic regression[J]. Ecological Modelling, 2000.
- [7] S.J.Back, K.M.Sung. "Fast K-nearest-neighbor search algorithm for nonparametric classification"[J]. Ctronics Letters,2000,36(21):1821-1822.
- [8] Yufei Tao, Dimitris Papadias, Nikos Mamoulis, Jun Zhang. "An efficient cost model for K-NN search technical report[J]. HKUST, 2001, 13(1): 1-14.
- [9] Bishop C M, Bishop C, Bishop C M, et al. Neural Network for Pattern Recognition. 1995.
- [10] Schmidhuber, Jürgen. Deep Learning in Neural Networks: An Overview[J]. Neural Netw, 2015, 61:85-117.
- [11] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, "Generative adversarial nets", Advances in Neural Information Processing Systems, 2014, pp.2672-2680.
- [12] Mogren O. C-RNN-GAN: Continuous recurrent neural networks with adversarial training. 2016.
- [13] Yin L, Zhang B. Time series generative adversarial network controller for long-term smart generation control of microgrids[J]. Applied Energy, 281.
- [14] Lin Z, Jain A, Wang C, et al. Using GANs for Sharing Networked Time Series Data: Challenges, Initial Promise, and Open Questions[C]// IMC '20: ACM Internet Measurement Conference. ACM, 2020.
- [15] Shuai L, Jin L, Min Z, et al. Adaptive Traffic Data Augmentation Using Generative Adversarial Networks for Optical Networks[C]// Optical Fiber Communication Conference. 2019.
- [16] Zhuang H, Zhao Y, X Yu, et al. Machine-Learning-based Alarm Prediction with GANs-based Self-Optimizing Data Augmentation in Large-Scale Optical Transport Networks[C], 2020 International Conference on Computing, Networking and Communications (ICNC). 2020.
- [17] Arjovsky M, Chintala S, L, et al. Wasserstein generative adversarial networks. 2017.
- [18] Cortez, P., Rio, M., Rocha, M., and Sousa, P. (2012). Multi-scale Internet traffic forecasting using neural networks and time series methods. Expert Systems, 29(2), 143-155.