

Analysis and Application Research of InfoGAN

Xinghao Wang*

Shanghai Jiao Tong University, Shanghai, China

*Corresponding author: kingsman@sjtu.edu.cn

Abstract. Generative Adversarial Networks (GAN) have been a revolutionary development in the field of Artificial Intelligence, particularly in the domain of generative models. The information maximizing generative adversarial nets, or infoGAN, is one of the most recent and promising developments in the world of GAN. InfoGAN focuses on maximizing the mutual information between the generator's output and some input variables. This means that the generated images or data can be controlled more easily while maintaining high-quality results. Apart from these applications, infoGAN has also been used in other areas such as natural language processing, anomaly detection, and even in music generation. With its versatility and robust performance, infoGAN looks set to become an increasingly important tool for researchers and practitioners in the field of machine learning. This paper focuses on the principle of infoGAN (information maximizing generative adversarial nets) and tries to put forward several ways to apply infoGAN to solve different kinds of problems in daily life.

Keywords: InfoGAN, GAN, Temperature field prediction, Fault diagnosis, STFT.

1. Introduction

The Generative Adversarial Network (GAN) network was introduced as a solution to the challenge of unsupervised learning, which has been a major obstacle for researchers in the field. In 2016, there were significant advancements in generative modeling that contributed to the success of unsupervised learning. Moreover, natural communication with machines is a dream goal and companies like Google and Facebook have proposed different approaches to achieve it [1]. However, there are still many challenges in natural language processing (NLP) that need to be addressed in order to reach this goal.

Unsupervised learning involves extracting patterns and structures from raw data without additional information, unlike supervised learning which requires labels. One of the classical methods for solving this problem using neural networks is the autoencoder [2]. The basic version consists of a multilayer perceptron where the input and output layers are of the same size, and a smaller hidden layer is trained to recover the input. The output of the hidden layer corresponds to representations of the data that can be used for clustering, dimensionality reduction, improved supervised classification, and data compression [2]. However, GANs were proposed in 2014 as a model that could solve unsupervised learning problems. Ian J. Goodfellow proposed the idea of GANs, which combines two networks: the generative network and the discriminator network [3]. The generative network is responsible for generating analog data, while the discriminator network determines whether input data is real or generated. The generated data is optimized constantly to reach a point where the discriminator network cannot differentiate between real and generated data. Meanwhile, the discriminator network is optimized to judge the data more accurately. The relationship between the two networks is antagonistic, hence the name adversarial network. Eventually, through contrarian learning between the generator and discriminator, a fake data set can be obtained that has almost the same distribution as the real dataset. Although the work is exceptional, there are still some issues within the GAN model, particularly with input noise. Therefore, the infoGAN can be considered a significant improvement over the GAN [4]. Since the input noise z of the generator is a continual noise signal without any kind of constriction, the GAN network can't take advantage of the noise z , which means that z can't be used as an interpretable representation. This is the problem infoGAN wants to solve. It tries to utilize the information of z to find out an interpretable representation.

In this case, the infoGAN disassembles noise z to two categories. One is the incompressible noise z and the other is the interpretable latent parameter c which is named as latent code. The infoGAN hopes to constrain the relation between c and the generated data set in order to help the latent code contain some kinds of interpretable information of the data set. For example, for the MNIST data set, the latent code c can be divided into categorical latent code which contains the information about the different types of digits and continuous latent code which can represent the inclination of digits and the thickness of stroke, etc.

2. Basic Idea

In order for the latent code c to be associated with the characteristic output of generated data, infoGAN introduces mutual information to constrain c . Since c has an interpretability for generated data $G(z, c)$, c and $G(z, c)$ should have a high correlation, that is, the mutual information between them is relatively high. Mutual information is a measure of the degree of dependence between two random variables [5]. The larger the mutual information is, the lower the information loss of latent code c will be when the generating network generates data according to the information of c . In other words, the more information of latent code c will be retained in the generated data. Therefore, infoGAN hopes that the bigger the mutual information between c and $G(z, c)$, the better. In this condition, the objective function of the model can be improved:

$$\min_G \max_D V_1(D, G) = V(D, G) - \lambda I(c : G(z, c)) \quad (1)$$

However, due to the difficulty in obtaining the true $P(c|x)$ when calculating the mutual information between c and $G(z, c)$, infoGAN uses the idea of variational inference by introducing variational distribution $Q(c|x)$ to approach $P(c|x)$. Thus the objective function of infoGAN has changed into:

$$\min_G \max_D V_{infoGAN}(D, G, Q) = V(D, G) - \lambda L1(G, Q) \quad (2)$$

As for the implementation part, the Q and D share all the convolution layers. Compared to the previous model, there is only an additional fully connected layer to output $Q(c|x)$. In this way, infoGAN does not increase the amount of calculation.

For the latent code c , if it is categorical latent code, it is possible to take advantage of the non-linear output of softmax to represent $Q(c|x)$ and if it is continuous latent code, it is possible can use the Gaussian distribution to represent $Q(c|x)$.

3. Fundamental Structure

3.1 Fundamental Structure of InfoGAN

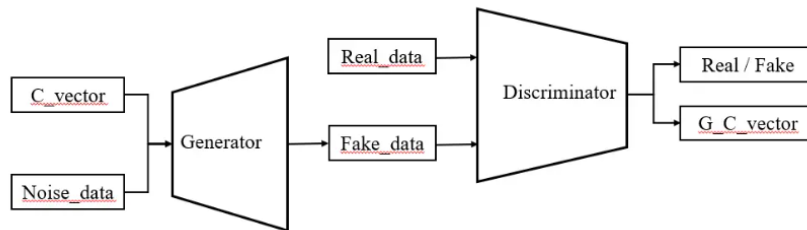


Fig. 1 Fundamental structure of InfoGAN

From the structure above, the Real_data part is used to mix with the generated Fake_data to participate in true or false judgement and updates the generator and discriminator based on the result of the judgement. In this process, the generate data can get closer and closer to the real data. The generated data not only needs to participate in the true or false judgement, but also needs to obtain mutual information with C_vector , and update the generator and discriminator according to the mutual information [5]. As a result, more information of C_vector is retained in the generated image. As shown in Figure 1.

3.2 Split Structure of InfoGAN

The basic structure of InfoGAN can be broken down as follows, with the discriminators D and Q sharing all the convolutional layers. The only difference lies in the final fully bonded layer. From another perspective, the common G-Q network is equivalent to a self-programmed network, G is equivalent to an encoder and Q is equivalent to a decoder, and the generated Fake_data is equivalent to the encoding of the input implicit variable C_vector . Hence, the divided structure of InfoGAN can be obtained [6]. As shown in Figure 2.

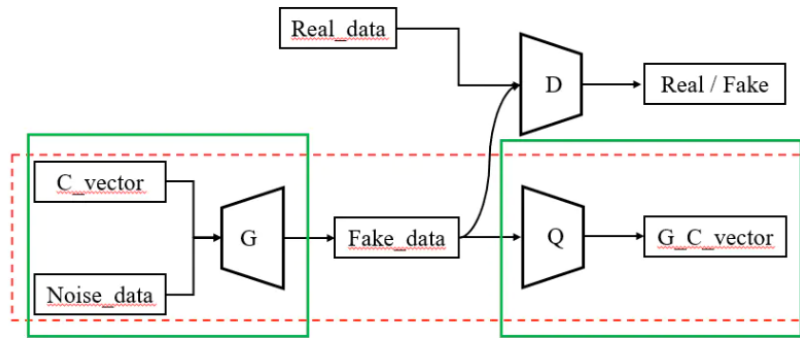


Fig. 2 Split structure of infoGAN

4. Experimental Result

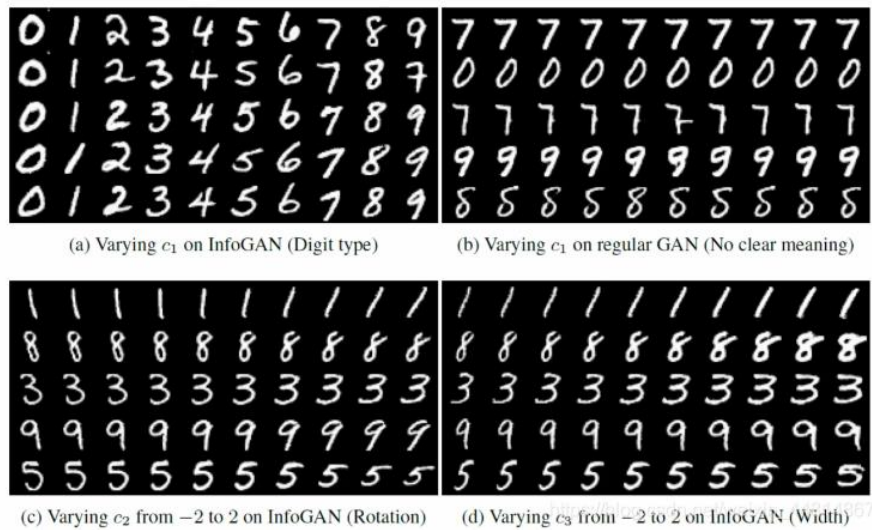


Fig. 3 Different latent code on InfoGAN(1)

From the result above, it is easy to tell that the latent code is quite important. Vector C_1 controls the exact type to generate; Vector C_2 controls the degree of inclination and vector C_3 controls the stroke thickness. A vertical comparison graph can make the result more explicit. As shown in Figure 3.

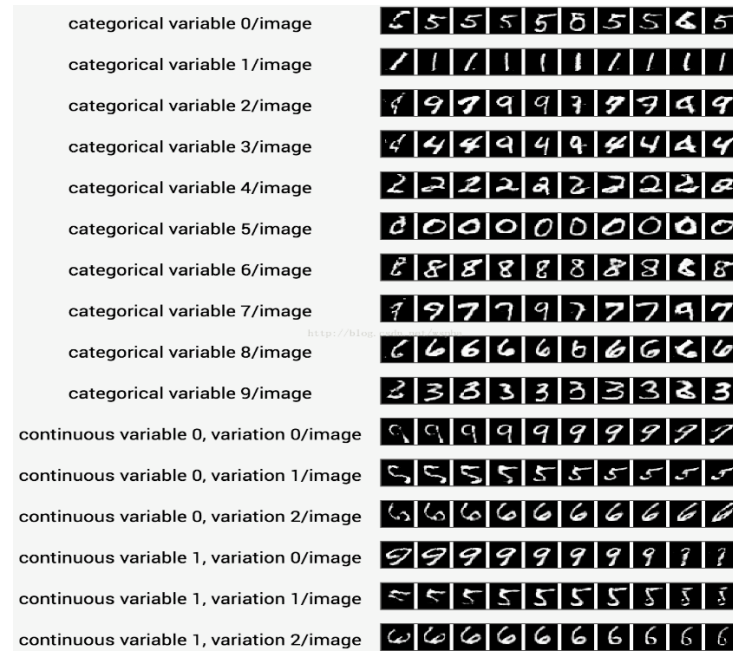


Fig. 4 Different latent code on InfoGAN(2)

From the graph here, modify the categorical latent code can generate different digital images and modify continuous latent code can change the degree of inclination and the thickness of strokes [7]. As shown in Figure 4.

5. Application Research

5.1 Bearing Small Sample Fault Diagnosis

5.1.1 Background

Generally speaking, bearings are one of the most important components of mechanical equipment. Due to the complex structure and working conditions, bearings are easy to be damaged and sometimes even cause injuries and deaths. In order to ensure safety, mechanical equipment is not allowed to operate under faulty conditions. In this condition, it is difficult to obtain a large number of bearing fault data sets [8]. In the process of bearing fault diagnosis, insufficient data or unbalanced data distribution are common.

5.1.2 Solution

InfoGAN network can be taken advantage of to solve the problem of small sample data. In this condition, one-dimensional vibration signal is used as the original information, and two-dimensional time-frequency images are generated after the pretreatment of short-time Fourier transform (STFT). These images can be used as the input of infoGAN to increase the training set and as a result, the model can be established [9]. Changing the latent code c can change the variation of the generated images. At the same time, there is a constraint between c and the generator, that is, the mutual information $I(c; G(z, c))$. In this case, this method based on the infoGAN can improve the quality of generated images and make the features of the data set more abundant.

5.2 Temperature Field Prediction Method for Industrial Heating Furnace

5.2.1 Background

Industrial heating furnace is an important equipment in refining and chemical plant. Whether it can operate safely directly affects the production and the economic benefit. In the process of operation, the furnace tube is often unevenly heated due to the poor combustion condition, and the phenomenon of local overtemperature happens. If the industrial heating furnace runs in the overtemperature state for a long time, it will eventually lead to coking, cracking, bulging and even tube explosion. Therefore,

measures must be taken to reduce the heterogeneity of heat intensity. However, the characteristics of transient change and random turbulence make it difficult to measure the temperature distribution, which leads to no reliable basis for combustion adjustment.

5.2.2 Solution

In this condition, using infoGAN can have a good benefit. The processed operating condition data can be regarded as latent code, and the generator can be trained to obtain a reconstructed temperature field which then can be put into the discriminator together with the CFD simulation temperature field for true or false judgment. As a result, the model can finally generate a reconstructed temperature field that is close enough to the CFD simulation temperature field based on the operating data.

5.3 Fault Diagnosis Method of Partial Discharge in Transformer

5.3.1 Background

As an important part of power system, the stability and safety of power transformer are directly related to the reliability of power supply. In oil-immersed power transformers, the structure of insulation system is complex. With the growth of service life, it is easy to appear insulation deterioration which leads to partial discharge inside the transformer [10]. Therefore, identifying the partial discharge type effectively is helpful to monitor and evaluate the insulation status of power transformers. In recent years, deep learning, represented by convolutional neural network (CNN), has become a popular research field in partial discharge pattern recognition. However, sufficient training samples with balanced distribution of categories are an important prerequisite to ensure the recognition effect of CNN-related network models. The problem is that the transformer failure state does not happen frequently and the probability of each fault type is different, which makes the field data scarce and unbalanced, leading to the lack of generalization ability of trained classifier.

5.3.2 Solution

Aiming at the problems of insufficient generalization ability of classifier and low recognition rate of general classifier in partial discharge fault diagnosis of transformer due to lack of data or imbalance between categories, a fault diagnosis method can be used for partial discharge of transformer based on improved residual network and InfoGAN. By simulating the gap types in oil-immersed power transformers, four types of PRPD spectra are collected as original samples. InfoGAN is used to expand the PRPD sample database, and its data enhancement effect was evaluated. Finally, the expanded sample data was put into the improved residual network for network training to obtain recognition results.

6. Conclusion

Undoubtedly, infoGAN represents a major breakthrough in the history of GANs. The concept of categorizing noise samples into two categories and obtaining an interpretable representation from the latent code is unprecedented. Moreover, there are several promising prospects for future research on infoGAN. Given that the mutual information is a key principle of infoGAN, it would be logical to extend the maximization of mutual information to other methods, such as VAEs. It may also be possible to learn hierarchical latent representations that could enhance semi-supervised learning and facilitate the discovery of high-dimensional data.

References

- [1] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (pp. 2672-2680).
- [2] Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein GAN. *arXiv preprint arXiv:1701.07875*.

- [3] Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1125-1134).
- [4] Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2018). Progressive growing of GANs for improved quality, stability, and variation. In Proceedings of the 6th international conference on learning representations.
- [5] Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.
- [6] Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232).
- [7] Brock, A., Donahue, J., & Simonyan, K. (2018). Large scale GAN training for high fidelity natural image synthesis. arXiv preprint arXiv:1809.11096.
- [8] Liu, Y., Zhang, X., & Ji, R. (2019). Diversity-promoting GAN: A cross-entropy based generative adversarial network for diverse text generation. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 3302-3312).
- [9] Mao, X., Li, Q., Xie, H., Lau, R. Y., Wang, Z., & Smolley, S. P. (2017). Least squares generative adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision (pp. 2794-2802).
- [10] Berthelot, D., Schumm, T., & Metz, L. (2017). BEGAN: boundary equilibrium generative adversarial networks. arXiv preprint arXiv:1703.10717