Isomorphic Wasserstein Generative Adversarial Network for Numeric Data Augmentation

Wei WANG¹,², Chuang WANG² and Yue LI¹,²,*

¹Key Laboratory for Medical Data Analysis and Statistical Research of Tianjin (KLMDASR), Tianjin 300350, China
²College of Computer Science, Nankai University, Tianjin 300350, China
*Corresponding author

Keywords: Isomorphism, Generative adversarial networks, Numeric data augmentation

Abstract. GAN-based schemes are one of the most popular methods designed for image generation. Some recent studies have suggested using GAN for numeric data augmentation that is to generate data for completing the imbalanced numeric data. Compared to the conventional oversampling methods, taken SMOTE as an example, the proposed GAN schemes fail to generate distinguishable augmentation result for classifiers. This paper introduces an isomorphic structure between generator $G$ and discriminator $D$ to the conventional WGAN, and hence develops an Isomorphic Wasserstein Generative Adversarial Networks (IWGAN). DGM-based analysis proves that the isomorphic structure establishes an additional restriction from $D$ to $G$ in learning $G$ and verse vice. Hence, the isomorphic structure enhances the classification performance in AUC on four datasets on five classifiers compared with three other GANs, and the conventional SMOTE methods add up to 20 groups of experiments. IWGAN outperforms all others in 15/20 groups.

Introduction

At present, multiple Generative Adversarial Network (GAN) schemes [1] have achieved significant progress in generating images and enhanced the accuracy of the classifier, where some of the GANs can produce almost indistinguishable images from human visional examination. In recent two years, several GAN models have been proposed for numeric data augmentation, which aims to generate samples to improve detection rates form multiple classifiers on the credit card fraud dataset [2, 3] and the telecom fraud dataset [4]. However, compared to the conventional augmentation methods, taken Synthetic Minority Over-Sampling Technique (SMOTE) [5] as an example, the GAN based methods have not exhibited many advantages [6].

Motivated by isomorphism in abstract algebra, we design an IWGAN for data argumentation. We define an isomorphic structure for the $G$ and $D$ pair. Here the isomorphic structure is defined as that the two networks have the same number of layers, each layer has the same number of nodes, and every two neighboring layers have the same connection. The two networks will be considered isomorphic or in same layers for the short of the definitions to satisfy requirements as mentioned above. Beneficial from the Wasserstein distance as the loss function, we technically setup the isomorphic network pairs, and the DGM analysis theoretically proves that this isomorphism provides an additional restriction in learning $G$ from $D$, and verse vice, respectively.

In evaluating of GAN-based augmentation, we compared IWGAN to three other GANs: conventional Wasserstein Generative Adversarial Network (WGAN) [6], adapted GAN proposed in 2017 [3], and GAN-DAE in 2018 [4]. In addition, the most widely used oversampling method, SMOTE [5], and is also employed in the evaluation as the baseline of data augmentation. Experiments are carried out on four widely studied datasets [8] and five classifiers, including Artificial Neural Network (ANN), Support Vector Machine (SVM), k-Nearest Neighbor (KNN), Gradient Boosting Classifier (GBC) and RF. In the common metrics, AUC in four datasets on five classifiers compared with three other GANs, and the conventional SMOTE methods add up to 20
groups of experiments. IWGAN outperforms all others in 15/20 groups. In the remaining three groups of experiments, the AUC index of IWGAN has one second best, two third best.

The remainder of the paper is organized as follows. We introduce an overview of previous related works on GAN and data augmentation in Section 2. We provide the proposed approach and the analysis through DGM in detail in Section 3. We show the further improved performance of IWGAN than SMOTE and other GANs using five classifiers on four representative datasets in Section 4. Finally, Section 5 presents the conclusions and outlines possible directions for future research.

**Related Works**

GAN is designed based on the idea of competition [10] the objective of the \( G \) is to confuse the \( D \). The \( D \) aims to distinguish the instances coming from the \( G \) and the instances coming from the original dataset. GAN is mainly used in the field of images to enhance the accuracy of the classifier [1]. As for the imbalanced numeric data, two main methods are under-sampling and over-sampling. Under-sampling balances the data set by reducing the size of the redundancy class. Conversely, over-sampling should use when the amount of data is insufficient. This method is to generate new samples by using SMOTE. SMOTE has been applied widely to the class imbalance problem. [5]

Recently, GAN has been used to generate samples to improve classifier performance in credit card fraud detection [2, 3]. Zheng et al. [4] adopted a deep denoising autoencoder to learn the complex probabilistic relationship among the input features effectively and employed adversarial learning that establishes a min-max game between a discriminator and a generator to accurately discriminate between positive samples and negative samples in the data distribution. Larsen et al. [11] presented an autoencoder that leverages learned representations to better measure similarities in data space. By combining a variational autoencoder with a generative adversarial network, it can use learned feature representations in the GAN discriminator as the basis for the VAE reconstruction objective.

The current GAN methods cause the unsatisfactory performance improvement of the classifier on the numeric datasets. The less effectiveness of GAN could be due to two factors, the dimensional differences and the representations in each dimension, as discussed in the introduction section.

**Proposed Method**

**GAN and WGAN**

GANs consists of two models, a generator model defined as \( G \) and a discriminator model defined as \( D \). \( G \) and \( D \) compete in a two-player min-max game. WGAN proposed by Arjovsky M et al. [6] completely solved the problem of GAN learning instability. WGAN uses Wasserstein distance (Earth-Mover) given by:

\[
W(P, P_g) = \inf_{\gamma \in \Pi(P, P_g)} \mathbb{E}_{(x, y) \sim \gamma}[d(x, y)]
\]

where \( \Pi(P, P_g) \) denotes the set of all joint distributions \( \gamma \sim (x, y) \) whose marginal are respectively \( P \) and \( P_g \). Kantorovich-Rubinstein duality:

\[
W(P, P_g) = \sup_{f \in \mathcal{F}} \mathbb{E}_{x \sim P}[f(x)] - \mathbb{E}_{x \sim P_g}[f(x)]
\]

\[
= \frac{1}{K} \sup_{f \in \mathcal{F}} \mathbb{E}_{x \sim P}[f(x)] - \mathbb{E}_{x \sim P_g}[f(x)]
\]

\[
\Rightarrow \max_{f \in \mathcal{F}} \mathbb{E}_{x \sim P}[f(x)] - \mathbb{E}_{x \sim P_g}[f(x)]
\]

The loss function of WGAN shown as follows:

\[
\min_{\theta} \max_{\theta'} E_{x \sim P}[f_{\theta}(x)] - E_{x \sim P_g}[f_{\theta'}(g_{\theta'}(z))]
\]
Compared to GAN, the problem of the collapse mode is almost solved by WGAN, ensuring the diversity of the generated samples.  

**IWGAN**

In order to design a stronger generator, we introduce an isomorphic structure, which is the same numbers of layer and the same layers in $G$ and $D$, so that a $G$-$D$ pair constraint can be added in the respective learning of $G$ and $D$. The challenge of setting up the same layer structure is the dimension differences of the inputs and outputs of $G$ and $D$, indicating different numbers of nodes in the first and last layer, between $G$ and $D$. Let $r^e$ represent the $d$-dimension of data $x$, $r$ represent $z$-dimension of noise $z$. Hence, we have the dimension of input for $G$ is $r^e$, and the input for $D$ is $r^e$ in GAN schemes. Benefited from Wasserstein distance as the loss function, the dimension of $D$’s output can be set various dimensions. Therefore, in IWGAN, $D$ output data in $r^e$, which is the same as the dimension of $G$’s output as shown in Figure 1. This Equation ensures the same last layer in $G$ and $D$.

![Figure 1. Comparison of discriminators’ structures of GAN, WGAN, IWGAN and The G’s structure of IWGAN.](image)

The problem of constructing isomorphic structure remains to establish the same input layer for $G$ and $D$. Generally, the dimension of $G$’s input $r$ is less than $r^e$, the dimension of the $D$’s input. The suggested mathematical is as follows.

Let $G$ and $D$ are functions in space, and we assume that space is Hilbert space. Firstly, generator $G$ is separated into two functions: one function, representing to the first layers corresponding the dimension conversion from $r$ to $r^e$, the other function is the following layers corresponding to multiple layer encoders. Then function $G$ forms a Hilbert space of a function, and

$$ G = g_i \cdot g_o $$  \hspace{2cm} (4)

The structure of $G$ is shown in Figure 1. It is reasonable to separate generator in mathematics, but the corresponding practical operation is still in challenge. The latest study proposed in 2019 [12] proved that in a sequence structure of a network, the different layers may have different weights. Hence we cannot be assigned the first several layers to function $g_i$, and the following layers to function $g_o$, while expecting the functions maintain the algebraically properties in Hilbert space.

Fortunately, the dimension of the numeric data is relatively low, where $R^e < 100$. Hence, we set $r^e = R^e$, even commonly $r^e < R^e$. Subsequently, we define an identical layer represented by $g$, as demonstrated in Figure 2 (b). We have $g_i = 1, g = 1 \cdot g_i$. Because of the existence of layer $1$, the distribution of return errors in different layers mentioned in [12] is not affected in the learning process. In this way, the decomposition method can be matched about designing networks in theory and practice. The advantage of this decomposition method is that in future research, even if $r^e$ is not equal to $r^e$, or $D$ has other specific structures, we can follow this line in further study. Through the decomposition of this function in space, the $G$-$D$ constraints between some sub-functions or all-functions are established and implemented in the network to generate a more effective GAN model.

**Discussion on Improvement of IWGAN**

This section discusses why and how establishing an isomorphic relationship between $D$ and $G$ can improve the performance of WGAN.
We employ the DGM to expound the working flow of WGAN and IWGAN. It is noticeable that in the actual network learning, we do not adopt the conditional probability and transition probability from DGM to optimize the models. We apply DGM to describe the learning mechanism and further evaluate the effectiveness of isomorphic structure in IWGAN.

In DGM representation, we define $G$, $D$, $E$ and $EG$ as random variables in Hilbert space respectively. $EG$ and $E$ are error distributions under the influence of multiple random variables. Then the whole process of WGAN can be represented by DGM in Figure 3 (a). Moreover, for $G$ and $D$ learning, we define the function $f$ to represent the learning process for $G$ and $D$ gave the observed random variables

$$f \left( G \mid D, E_G, E \right)$$

for learning $G$ 

$$f \left( D \mid G, E_G, E \right)$$

for learning $D$ 

After factorization in DGM, there are

$$f \left( G \mid D, E_G, E \right) = f \left( G \mid E_G, D \right)$$

(7)

$$f \left( D \mid G, E_G, E \right) = f \left( D \mid E \right) f \left( D \mid E_G, G \right)$$

(8)

The above description conforms to the detail learning process of WGAN. When learning $G$ via Equation (7), we calculate $E_G$ to minimize $G$ under given $D$. In addition, when learning $D$ from Equation (8), besides the constraints of $E$, we calculate $E_G$ to minimize $D$ under given $G$, which satisfies the adversarial relationship between $G$ and $D$.

At first, in the isomorphic structure, we define an isomorphic function that acts on the $G \to D$ Hilbert space given by:

$$D = TG$$

(9)

Correspondingly, in DGM we introduce a hidden variable $T$, and the discriminator function becomes the $TG$. It is noticed that we do not change the learning process of WGAN. So we still try to solve the optimal $D = TG$, not hidden variable $T$ when learning stage for the discriminator.

By introducing $T$, we can see that two flows from $T \to D$ and $G \to D$ are added in DGM. Then the whole process of IWGAN can be represented by Figure 3 (b).

In the process of learning, the learning function becomes

$$f \left( G \mid T, TG, E_G, E \right) = f \left( G \mid E_G, TG \right) f \left( G \mid TG \right)$$

$$= f \left( G \mid E_G, D \right) f \left( G \mid D \right)$$

(10)
\[ f(D | T, G, E_0, E) = f(T | G) \cdot f(G | E_0, G) \cdot f(T | G) \\
= f(D | E) \cdot f(D | E_0, G) \cdot f(D | G) \]

Due to the existence of \( T \rightarrow T G \), \( f(G | D) \) for learning \( G \) and \( f(D | G) \) for learning \( D \) contain an extra constraint of \( D \leftrightarrow G \), which are \( f(G | D) \) in Equation (10) and \( f(D | G) \) in Equation (11) added in the learning process of DGM. From a certain point of view, the mutual constraint of \( D \leftrightarrow G \) explains why adding an isomorphic construction can generate stronger data for classifiers, improve the classification performances.

![Diagram](image.png)

**Figure 3.** The processes of (a) WGAN and (b) IWGAN represented by DGM.

### Experiments and Results

#### Experiments

The evaluation study is directed to determine the influence of the performance of IWGAN, GANs and SMOTE, using a variety of classifiers on four Datasets. According to this aim, experiments are conducted as follows:

**Datasets:** We conduct experimental analyses based on four datasets, which are obtained from the University of California Irvine (UCI) machine learning repository [7]. The four binary classification datasets with various imbalance ratios are Australian Credit Approval data, German Credit data, Pima Indians Diabetes data and SPECT heart data.

**Classifiers:** Five general classifiers, Artificial Neural Network (ANN), Support Vector Machine (SVM), k-Nearest Neighbor (KNN), Random Forest Classifier (RF) and Gradient Boosting Classifier (GBC) are applied with the default parameter settings, recommended by scikit-learn [13].

**Evaluated Methods:** We compare the proposed IWGAN with the state-of-the-arts in augmenting data by GANs, including adapted GAN proposed in 2017 [3], GAN-DAE in 2018(Zheng, 2018), and the conventional WGAN under the baseline provided by SMOTE. Among these methods, WGAN without isomorphic structure request is used as another baseline for IWGAN.

**Evaluation metrics:** In this paper, AUC, as recommended in [9, 14, 15], are calculated as appropriate evaluation criterion to measure the augmentation performance in classification.

Figure 4 demonstrates the AUC in four datasets on five classifiers compared with three other GANs, and the conventional SMOTE methods add up to 20 groups of experiments. IWGAN outperforms all others in 15/20 groups. In the remaining three groups of experiments, the AUC index of IWGAN has one second best, two third best. Among the datasets, SPECT dataset is the least sensitive to the classifier and augmentation methods, while the other three datasets demonstrate the distinguishable result. About classifiers, GBC outputs the relatively best results on all of the datasets, while KNN produces the worsts. Affected by the datasets and classifiers, GAN-DAE and GAN generate unstable augmented data. For example, on all of the datasets classified by RFC, the data which is generated by GAN-DAE and GAN has a worse result than data generated via SMOTE. However, when the datasets classified by SVM, GAN-DAE and GAN generated data illustrate better results than SMOTE in three of the four datasets. In our experiments, IWGAN outperforms all other GANs in all test cases except RF in Pima dataset. Meanwhile, IWGAN also outperforms the baseline, which is SOMTE, except RF in Pima dataset and SVM in SPECT dataset.
Conclusion and Future Extensions

This paper proposed IWGAN based on the isomorphic structure to data augmentation on four UCI publicly datasets. In addition, we technically setup the isomorphic network pairs, and the DGM analysis theoretically proves that this isomorphism provides an additional restriction in learning G from D, and verse vice, respectively. Moreover, in the common metrics, AUC in four datasets on five classifiers compared with three other GANs, and the conventional SMOTE methods add up to 20 groups of experiments. IWGAN outperforms all others in 15/20 groups. In the remaining three groups of experiments, the AUC index of IWGAN has one second best, two third best.

Figure 4. Five data augmentation methods of AUC for five classifiers in (a) Australian Credit Approval dataset (b) German Credit dataset (c) Pima Indians Diabetes dataset (d) SPECT Heart dataset.

Subsequently, some isomorphic functions may exist even if G and D do not satisfy the requirements of the numbers of layers and the same layers. Technically, we want to set up different IWGANs with relative isomorphic structures, and find that the smaller the difference between the number of nodes in G and D, the higher the effect in the future. And we further study the isomorphism of GAN partial layers in other cases, to improve their performance on numeric data augmentation and even image generation.

Acknowledgement

This work was supported by the Tianjin Natural Science Foundations(17JCYBJC23000), National Key Research and Development Program of China (2016YFB0201304), Fundamental Research Funds for the Centeal Universities(Nankai University 070/63191114)and National Key Research and Development Program of China under the grant number(2018hjyzkfkt-002)

References


