重庆师范大学

全日制本科生毕业设计

题　目：生成对抗网络模型的泛化和均衡研究

学　　院：计算机与信息科学学院

专业年级：计算机科学与技术 2016 级

学生姓名：云伟东　　学号：2016051603202

指导教师：曾　智　　职称：教　授

2020 年 04 月 30 日

生成对抗网络模型的泛化与均衡研究

计算机信息与科学学院　计算机科学与技术专业　2016级　云伟东

指导教师　曾智

摘　要：在互联网时代快速发展的当代，各式各样的数据不断地涌现出来。人们开始考虑如何从大数据中获得对这个世界的认知。生成对抗网络（GAN）以其独特的对抗性训练方式和内涵的机器博弈思维照亮了人工智能发展道路。当然，GAN在训练过程中无法稳定和全局收敛。在众多GAN模型中比较经典的Wasserstein GAN改进了度量数据分布的距离，抛弃了JS散度和KL散度，EM距离的优势特别明显，同时以其独特的目标函数惩罚机制使得训练过程比较稳定。当然在目标函数的角度下，LS-GAN（损失敏感）和GLS-GAN（广义的LS-GAN）的“按需分配的能力”，充分发挥了Lipschitz属性。同时从梯度向量角度来稳定训练过程，零梯度惩罚方法和本文的梯度中心化算法是直接作用在梯度向量上使得GAN的泛化和平衡得到提高。本文将要从目标函数、梯度向量优化和博弈论思维的角度来尝试探究GAN模型训练过程中要达到的纳什均衡状态。

关键词：生成对抗网络；纳什均衡；对抗训练；博弈论思维

Abstract：In the rapid development of the Internet, all kinds of data are constantly emerging. People are starting to think about how they can learn about the world from big data. Generative antagonism network (GAN) illuminates the development path of artificial intelligence with its unique training method of antagonism and machine game thinking. Of course, GAN cannot stabilize and converge globally during training.Among many GAN models, Wasserstein GAN, which is more classical, improves the distance of measurement data distribution, and abandons JS divergence and KL divergence. The advantage of EM distance is particularly obvious. Meanwhile, it makes the training process more stable with its unique penalty mechanism of target function.Of course, from the perspective of the objective function, the "ability to allocate according to demand" of LS-GAN (loss sensitivity) and GLS-GAN (generalized LS-GAN) gives

full play to the Lipschitz attribute.Meanwhile, from the perspective of gradient vector to stabilize the training process, the zero gradient penalty method and the gradient centralization algorithm in this paper directly act on the gradient vector to improve the generalization and equilibrium of GAN. This paper attempts to explore the Nash equilibrium state to be achieved in the training of GAN model from the perspectives of objective function, gradient vector optimization and game theory.

Key words：Generative antagonistic network; Nash equilibrium; adversarial training; Game theory thinking

1 生成对抗网络（GAN）的纳什均衡研究背景

生成对抗网络自 2014 年被提出以来，经过近几年的快速发展，在非结构化数据领域取得了突出的效果。然而其训练方式导致的问题使得它显得缺陷很大。无数学者都在 GAN 基本模型上进行优化改进，例如目标函数角度、梯度惩罚和应用场景等等。

那么纳什均衡状态是何种状态？纳什均衡状态就是生成器和判别器相互对抗博弈，最终到达一种无法再通过调整策略增加收益的状态。纳什均衡状态的达到与否直接关系着模型收敛的好坏。由于在数学理论上，理想的纳什均衡似乎永远达不到，只能退而求其次，寻找预期的局部最优点均衡。而这局部最优点纳什均衡一种方式则是通过提高 GAN 模型中判别器的泛化能力所希望达到的一种均衡状态。

经过国内外学者长时间地对 GAN 模型训练地数学理论上的研究因为涉及了博弈论、动力学和势场等学科领域而进展缓慢[1]。然而研究者对神经网络模型的泛化能力和对抗样本攻击研究的突破，使得将神经网络模型的泛化和纳什均衡联系在一起有着光明的前景。这使得在数学理论等交叉性学科比较薄弱的人工智能研究员们目光吸引在这个领域。虽说，研究角度发生了改变，但是最终还是要解

决关于GAN模型的几大根本性问题。以下，将逐步介绍开放性问题及理论原理。

2 GAN模型的几大开放性问题

2.1 GAN与其他生成模型的权衡是什么？

现在人工智能领域较为成功的生成模型有生成对抗网络、流模型和自回归模型。自回归模型使用的损失函数是Pixel Loss（像素损失），通过实验表明自回归模型效率较低且生成的图像质量不行。流模型也因为最大似然训练比对抗训练更加难计算导致流模型的效率较低。但是流模型和自回归模型仍然有着GAN所没有的一些优势。常见的图像损失如表2.1所示。

表2.1 常见的图像损失

图像损失	优点	缺点
Pixel loss	使用简单，训练速度快，稳定	输入图像模糊，质量较低
GAN loss	提高生成图像质量，更加真实	学习整体生成分布，无法单独使用
Perceptual loss	注重图像包含的高维特征	受限于预训练的其他神经网络

2.2 GAN可以模拟哪种分布？

GAN的大多数研究是在图像合成领域，例如CycleGAN、ConditationGAN和StyleGAN等。所以，学者大都在少数开源标准数据集上进行训练GAN模型。GAN模型中的生成器和判别器的内部结构也大都是神经网络结构，那么GAN是如何通过神经网络拟合一个图像数据分布的呢？是否存在GAN永远无法学习拟合的图像数据分布？

2.3 除了图像合成领域，GAN还适合哪些领域？

2.3.1 文本数据的离散性问题

对于文本数据研究比较深入就是循环神经网络和LSTM（长短时记忆网络）了，而GAN是通过将来自鉴别器的梯度信息反向传播到生成器中以此来促进训练。虽然生成器和判别器的内部结构也可以是循环神经网络，但是文本数据的的

离散性带来的问题是比较难解决的。而现在在文本数据上应用较为广泛的预训练模型参数极多使得在 GAN 模型上对抗训练较为困难。

2.3.2 GAN 在结构化数据和非结构化数据（如图形）上的应用前景

深度学习对于非结构化数据的研究在当今的大数据和强大算力的加持下达到了深不可测的高度。对于结构化数据（如知识图谱等）深度学习就显得乏力了，但是最近比较火的图神经网络（GCN）比较适合那些非欧式结构化数据。

2.3.3 音频领域

音频领域的研究复杂程度在于它涉及了很多领域：如语音识别、自然语言处理和情感分析等等。其在预处理方面也是比较严格，如果预处理方面比较差的话会导致后续的数据研究的精度极差。它涉及了传统机器学习、神经网络和声音声波等专业的领域，如果仅仅依靠 GAN 是无法完全掌控的。

2.4 GAN 是如何保证在训练过程中收敛的？

训练 GAN 与训练其他神经网络不同，因为 GAN 模型的训练过程中是使用完全对立的目标函数优化了生成器和鉴别器。在某些假设下这种迭代异步优化是局部渐近稳定的。因为鉴别器和发生器的损失函数是非凸函数，所以无法证明 GAN 是否可以在全局训练过程中收敛。那么怎么证明 GAN 在全局收敛呢？答案是博弈论。只要从对抗训练的本质出发，才能证明是否全局收敛。但是博弈论所研究的纳什均衡是一种理想状态下的均衡，GAN 模型在训练过程中只能达到局部的鞍点。目前的研究方法大都是尽量使得训练过程达到预期的局部最优均衡点，以此来稳定训练。

2.5 GAN 与对抗样本有什么联系？

众所周知，图像分类器遭受对抗样本的困扰：人眼无法察觉的轻微像素级别的扰动会导致分类器对输入图像的类别给出天地之别的输出。尽管有关 GAN 和对抗样本的研究成果很多，但人们对于它们之间的关系研究较少且应用也寥寥。那么对抗样本影响下的判别器网络的鲁棒性高低对 GAN 的训练产生何种影响?判别器对于对抗样本的敏感与否直接关乎着泛化能力的高低。如果只是在图像的像素

空间中对像素极为敏感，那么模型对于图像的纹理和特征的学习就显得极其薄弱。这样的人工智能只是像素值的记忆机器罢了，根本没有形成类似人类思维的概念。

3 Wasserstein GAN、LS-GAN 以及 GLS-GAN

3.1 从正则项了解如何让 GAN 收敛

首先将使用一个简化形式的 GAN 模型作为测试对象，训练数据的位置固定在 x=0；生成器的生成样本的位置在 x= θ ，如图 3-1 所示：

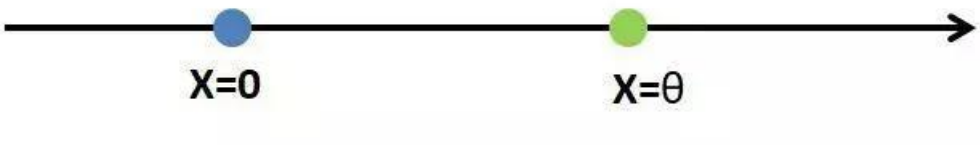


图 3-1 一维空间下数据样本点

判别器为一个简单的线性函数与激活函数复合的形式如

$$D_{\varphi}(x)=\mathrm{f}(\varphi\bullet x) \tag{3-1}$$

激活函数使用 Sigmoid 函数

$$\mathrm{f}(t)=\sigma=\frac{1}{1+e^{-t}} \tag{3-2}$$

可获得原始形式

$$f(t)=t \tag{3-3}$$

简化形式的 GAN 的纳什均衡点为(0,0)，即生成的样本与训练数据重合。如图 3-2 所示。

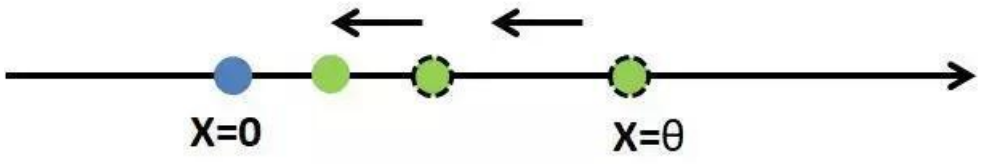


图 3-2 对抗训练下数据样本点拟合图

之后，在众多不同的 GAN 模型能否收敛到均衡点呢？实际情况远远比简化 GAN 更加复杂和更加涉及到多维空间，样本不仅不能只存在于一个维度，也不能存在于一个抽样点，通过这个点可以看到一些东西并得到一些启示。

标准 GAN 即 Ian Goodfellow 首次提出的 GAN 的标准形式，其损失函数的表达式[2]

$$\max_{D} \mathrm{E}_{x\sim pdata}[\log D(x)]+\mathrm{E}_{x\sim pg}[\log(1-D(x))] \tag{3-4}$$

$$\min_{G} \mathrm{E}_{z\sim pz}[\log(1-D(G(z)))] \tag{3-5}$$

在简化 GAN 中，对应的损失函数

$$\max_{\varphi} \log[1-\sigma(\varphi\theta) \tag{3-6}$$

$$\min_{\theta} \log[1-\sigma(\varphi\theta)] \tag{3-7}$$

相应的动力学系统

$$\begin{pmatrix} \dot{\varphi} \\ \dot{\theta} \end{pmatrix} = \begin{pmatrix} \theta\,\sigma(\varphi\theta) \\ -\varphi\,\sigma(\varphi\theta) \end{pmatrix} \tag{3-8}$$

采用梯度下降法发现其并不收敛，其情况如图 3-3 所示：

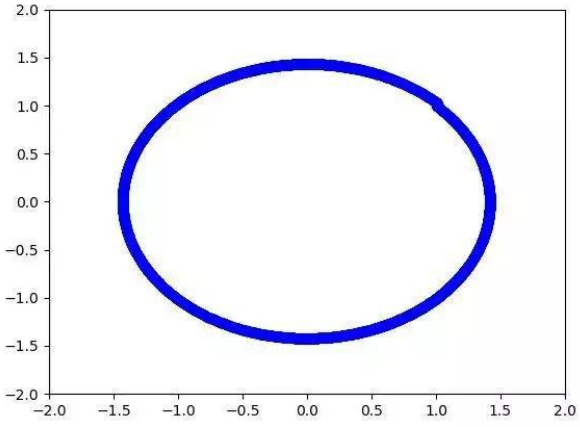


图 3-3　对抗训练下梯度下降图

3.2　标准 GAN 和其“无限的建模能力”[3]

学者研究表明[4]，GAN 是一种通过对输入的随机噪声 z（比如高斯分布或者均匀分布），通过神经网络 G 拟合成一个新样本，该样本的分布希望和真实数据的分布类别一致（比如图像、视频等）。

Ian Goodfellow 假设用于评估数据分布真实性的 Discriminator 具有无限的建模能力，这意味着无论实际数据分布和生成数据分布多么复杂，D 网络都可以将其分开。这个假设叫做非参数假设。对于深度神经网络来说，只要不断地加宽加深，就可以达到预想到的结果。

当分布之间的真实数据和生成样本交点微不足道，而又由于 D 网络具有对数据分布差异可以完美分开的能力。经典 GAN 用来度量生成数据分布和真实数据分布的相识度（JS 散度）就会变成一个常数！深度学习算法基本是用梯度下降法来优化网络的。一旦 JS 散度为常数，其梯度就会消失，也就会使得无法对 G-网络进行持续地更新，训练过程就停止了。

3.3 WassersteinGAN 和解决梯度消失问题的方法

WGAN 提出了取代 JS 散度的 Wasserstein 距离来测量真实和生成样本密度之间的距离。这个距离的特点是，即使完美地分割真实的数据样本并创造无限可能的数据样本，距离也不会退化为常数，并且仍然可以提供一个梯度来优化 G 网络。如下定义真实分布与生成分布的 Wasserstein 距离[5]

$$W(P_r,P_g)=\inf_{\gamma\sim\prod(P_r,P_g)}E_{(x,y)}[\|x-y\|] \tag{3-9}$$

其中，Pr 和 Pg 分别为真实分布与生成分布，γ 为Pr 和 Pg 的联合分布。相较于 JS 散度和 KL 散度，Wasserstein 距离的好处是即使它们不相交，Wasserstein 距离仍然可以对两个分布式距离做出清晰的反应。为了与 GAN 相结合，将其转换成对偶形式

$$W(P_r,P_g)=\sup_{\|f\|_L\le 1}(E_{x\sim P_r}f_w(x)-E_{x\sim P_g}f_w(x)) \tag{3-10}$$

不同之处在于 WGAN 不再需要将鉴别器分类为 0-1，将其值限制在(0.1)之间，这意味着 fw 越大，它离实际分布越近；相反，它离分配越近。此外，$\|f\|_L\le 1$表

示其 Lipschitz 常数为 1。Lipschitz 显然很难在判别器上是连续约束的，以便更好地表达 Lipschitz 转化成权重剪枝，即要求参数 w ∈ [-c, c]，其中为常数。因而判别器的目标函数

$$\max_{f_w} E_{x\sim P_r}[f_w(x)] - E_{z\sim p_z}[f_w(G(z))] \tag{3-11}$$

其中 w ∈ [-c, c],生成器的损失函数

$$\min_{G} -E_{z\sim p_z}[f_w(G(z))] \tag{3-12}$$

WGAN 理论上解释了了因生成器梯度消失而导致训练不稳定的原因，并用 Wasserstein 距离替代了 JS 散度，解决了梯度消失问题。继而，为了改进与实际分布相关的距离理论，以及从 Wasserstein 的距离产生的 Lipschitz 限制，也给后人带来了更深层次的灵感，如基于 Lipschitz 密度的损失敏感 GAN(loss sensitive GAN, LS-GAN)。

WGAN 的优化目标

$$f_w^* = \arg\max_{f_w\in\Gamma_1} \cup(f_w, g_\phi^*) \overset{\Delta}{=} E_{x\sim Pdata}[f_w(x)] - E_{z\sim Pz(z)}[f_w(g_\phi^*(z))]$$
$$and \tag{3-13}$$
$$g_\phi^* = \arg\max V(f_w^*, g_\phi) \overset{\Delta}{=} E_{z\sim Pz(z)}[f_w^*(g_\phi(z))]$$

WGAN 的损失函数图如 3-4 所示：

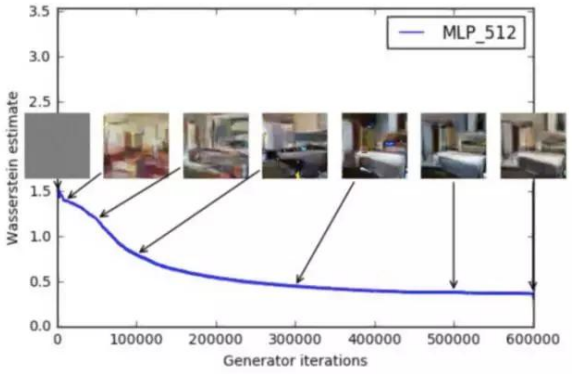


图 3-4　WGAN 的损失函数图[6]

4　梯度角度：Gradient Centralization（梯度中心化算法）

自从深度学习逐渐火爆起来，学者们不断推出新的优化技术以此来达到提高训练速度和减轻在训练数据集上的过拟合程度。但是现在的大多数优化技术比如权重标准化、批归一化和实例归一化等等对于激活函数或者权重向量进行改进思路。研究者另辟蹊径直接对梯度下手，提出全新的梯度中心化方法。通过对梯度中心化算法一段时间的特性和有点的研究和实践，通过实验结果发现梯度中心化算法（GC）在 WGAN、GLS-GAN 等模型上使用会产生一些意想不到的效果，例如使得 GAN 模型训练更加的稳定、生成图像的质量和多样性得到了大幅度的提高等等。

4.1　Gradient Centralization（梯度中心化算法）的原理

梯度集中算法（GC）通过将梯度向量分布的均值设置为 0 直接对梯度进行操作[7]。同时 GC 具有 Lipschitz 约束损失函数特性的投影梯度下降法使得该优化算法在高维流形空间对于拟合数据分布在多维梯度上都有着实质性的改进，因此训练过程变得更加有效和稳定。其具体情况图 4-1 表示：

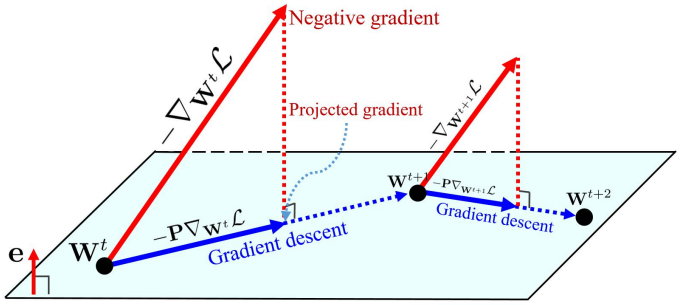


图 4-1　投影梯度下降图[8]

众所周知，标准的 GAN 模型在训练过程中当 G 和 D 两者的 Capacity 足够时，模型会收敛，二者达到无法再通过调整策略改进自己受益函数的纳什均衡状态。此时 $P_{data}(x)=P_g(x)$，判别器不论是对于 $P_{data}(x)$ 还是 $P_g(x)$ 中采样的样本，其预测概率均为 1/2，表示生成样本与真实样本达到了难以区分的地步。之后，通过一系列的数学推理，得出 GAN 的目标函数是和 KL 散度和 JS 散度有关系的。

众所周知，创建一个对抗网络模型是为了优化 KL 的分散性和 JS 的分散性。根据 JS 分散特性，当两个分配间隔不相交时，它们的值趋于 Log2 常数，这就是梯度消失的原因。这意味着下降的梯度是零。当接近最佳鉴别器时，生成器肯定不会得到任何梯度信息;在博弈论思维下，博弈的进行的要求在与不断有着信息在博弈的两者之间流动,在信息的促进下博弈才能接近那理想状态下的纳什均衡状态。

以上可知在（近似）最优判别器下，最小化生成器的损失函数等价于最小化 P_r 与 P_g 之间的 JS 散度，而由于 P_r 与 $p_{\sim}$ 几乎不可能有不可忽略的重叠，所以这种情况下 JS 散度都是常数 log2，最终导致生成器的梯度（近似）为 0。

根据 WGAN 的论点，生成器梯度已经消失，数据分布不能只在二维空间中看到。与此同时，Wasserstein 距离的提出使得在梯度优化 KL 和 JS 散度方面改善了突变。那么如果能够把 Wasserstein 距离定义为损失函数，就可以产生有意义的梯度来更新生成器，使得生成分布逐渐较好地拟合真实数据分布。

WGAN 模型在模型训练过程中的成功也启发了研究 GAN 模型的众多学者，经过近几年不断的改进和优化出现了很多比 WGAN 更好的 GAN 模型。例如：限定 GAN 模型建模能力的 LS-GAN（损失敏感）、对 LS-GAN 进行有监督和半监督推广的 GLS-GAN、改进 Wasserstein 距离的 Banach Wasserstein GAN、Spectral Norm Regularization（谱归一化）和由两个时标更新规则（TuTuR 规则）训练的 GAN 收敛到局部 Nash 平衡等等。

但是这些大部分研究的角度仅仅限于目标函数的优化，通过目标函数的优化间接地去影响梯度下降优化。那么有没有一种直接进行梯度下降优化的算法呢？梯度中心化算法就是通过设置梯向量分布的均值为零来中心化梯度向量，使学习过程更加有效和稳定。判别器训练是稳定的，它给生成器的梯度会更稳定，从而实现稳定模型的训练目标。

4.1.1　梯度中心化算法研究目的

通过实验发现 GAN 模型中的 D（判别器）的泛化能力较低会导致“模式坍塌”的现象出现。

同时 Arora 等人(2017)证明了生成器可以通过记住一个多项式数量的训练实例来获胜。结果表明，低容量鉴别器检测不出多样性的缺失。因此，它不能教发生器接近目标分布。但是经过实验发现，使用原始 GAN 损耗训练的高容量鉴别器倾向于对训练数据集中错误标记的样本进行过度拟合，从而引导生成器走向崩溃均衡(即生成器有模式崩溃的均衡)。或许只要提高 GAN 判别器的的泛化能力减少过拟合程度会引导生成器不会走向崩溃均衡。过拟合判别器不是将模型数据分布向目标分布靠近，而是过度的靠近数据集中的真实样本。这就解释了为什么原始 GAN 通常表现出模态崩溃行为。之后，利用梯度下降法寻找经验最优判别器（与最优判别器不同），通常需要多次的迭代。但是 GC 能从提升损失函数和梯度两种角度的 Lipschitz 属性，使得梯度下降中的梯度不易发生突变，从而使训练过程更加高效和稳定。

那么为什么 GAN 的判别器的泛化能力会影响模型训练，导致“模式坍塌”的问题出现？或许可以从 L 约束和泛化能力的角度进行解释。

记输入为 x，输出为 y，模型为 f，模型参数为 w，其参数函数

$$\mathrm{y} = f\{w\}(x) \tag{4-1}$$

在改进 GAN 模型的性能时通常希望得到一个稳定训练的模型。对于参数扰动的稳定性和输入扰动的稳定性达到预期效果下的较小模型就可以稳定训练。经过研究表明深度学习模型存在“对抗攻击样本”，比如图片只改变某个像素就给出完全不一样的分类结果，这就是模型对输入过于敏感，泛化能力就比较小。

也就是说，希望 ||x1-x2|| 很小时

$$\|F_w(x_1) - F_w(x_2)\| \tag{4-2}$$

也尽可能地小。当然，“尽可能”究竟是怎样，谁也说不准。于是 Lipschitz 提出了一个更具体的约束，那就是存在某个常数 C（它只与参数有关，与输入无关），使得

$$\|F_w(x_1)-F_w(x_2)\|\le c(w)\bullet\|x_1-x_2\| \tag{4-3}$$

公式 4-3 中 c(x)越小意味着它对输入扰动越不敏感，泛化性越好。

同时，如果 GAN 模型中的 D（判别器）的泛化能力不高，对输入样本比较敏感会导致 G（生成器）的生成样本更加向真实样本的多峰数据分布去拟合，继而会导致 G 生成样本比较单一，多样性低，导致模式坍塌的问题出现。

这种现象可以很好的由如图 4-2 所示：

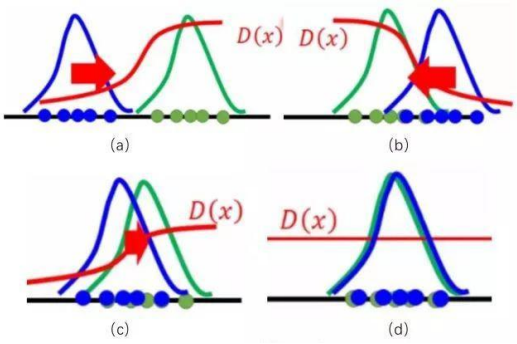


图 4-2　对抗训练下生成数据分布与真实数据分布拟合图

因此梯度中心算法可以很好地避免学习过程中的梯度爆炸，增加了 Lipschitz 约束属性的 D 来限制损失。

4.1.2　梯度中心化算法基本原理

在模型训练过程中生成器拟合噪声数据分布和判别器判断数据分布来源，梯度中心化算法核心公式

$$\Phi_{GC}(\nabla_{W_i}\ell)=\nabla_{\mathrm{w_i}}\ell-\mu\nabla_{\mathrm{w_i}}\ell \tag{4-4}$$

其中

$$\mu\nabla_{w_i}\ell=\frac{1}{m}\sum_{j=1}^{m}\nabla_{w_{i,j}}\ell \tag{4-5}$$

公式 4-4 的矩阵表述

$$\Phi_{GC}(\nabla_W\ell)=P\nabla_W\ell,\ \mathrm{and}P=I-\mathrm{ee}^T \tag{4-6}$$

如图 4-3 展示了通过梯度中心化算法改进优化算法 SGDM 和 Adam 的具体情况 。此外，如要使用权重衰减，可以设置

$$\hat{\mathbf{g}}^{t} = P(\mathbf{g}^{t} + \lambda w) \tag{4-7}$$

其中 λ 表示权重衰减因子。如图 4-3 所示。

Algorithm 1 SGDM with Gradient Centralization

Input: Weight vector $\mathbf{w}^0$, step size α, momentum factor β, $\mathbf{m}^0$

Training step:

1: **for** $t = 1, \ldots T$ **do**
2: $\mathbf{g}^t = \nabla_{\mathbf{w}^t}\mathcal{L}$
3: $\widehat{\mathbf{g}}^t = \Phi_{GC}(\mathbf{g}^t)$
4: $\mathbf{m}^t = \beta\mathbf{m}^{t-1} + (1-\beta)\widehat{\mathbf{g}}^t$
5: $\mathbf{w}^{t+1} = \mathbf{w}^t - \alpha\mathbf{m}_t$
6: **end for**

Algorithm 2 Adam with Gradient Centralization

Input: Weight vector $\mathbf{w}^0$, step size α, β_1, β_2, ϵ, $\mathbf{m}^0$,$\mathbf{v}^0$

Training step:

1: **for** $t = 1, \ldots T$ **do**
2: $\mathbf{g}^t = \nabla_{\mathbf{w}^t}\mathcal{L}$
3: $\widehat{\mathbf{g}}^t = \Phi_{GC}(\mathbf{g}^t)$
4: $\mathbf{m}^t = \beta_1\mathbf{m}^{t-1} + (1-\beta_1)\widehat{\mathbf{g}}^t$
5: $\mathbf{v}^t = \beta_2\mathbf{v}^{t-1} + (1-\beta_2)\widehat{\mathbf{g}}^t \odot \widehat{\mathbf{g}}^t$
6: $\widehat{\mathbf{m}}^t = \mathbf{m}^t/(1-(\beta_1)^t)$
7: $\widehat{\mathbf{v}}^t = \mathbf{v}^t/(1-(\beta_2)^t)$
8: $\mathbf{w}^{t+1} = \mathbf{w}^t - \alpha\frac{\widehat{\mathbf{m}}^t}{\sqrt{\widehat{\mathbf{v}}^t}+\epsilon}$
9: **end for**

图 4-3　梯度中心化算法的算法流程图[9]

下图展示了四种组合的训练损失和测试准确率曲线。如图 4-4 显示，使用单一变量对比损失函数和精度曲线的变化趋势，梯度中心化算法在训练速度和泛化能力两方面都有着较好的改进空间。其如图 4-4 所示：

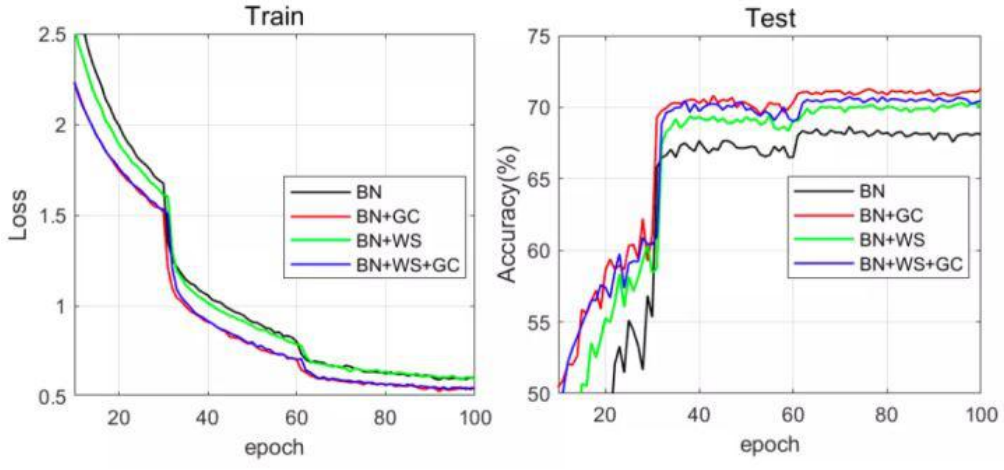


图 4-4　梯度中心化算法在数据集上训练对比图[10]

4.1.3　梯度中心化算法在 WGAN、LS-GAN 和 GLS-GAN 上的应用

梯度中心化算法优化技术以其在常见的开源数据集（ImageNet 等）和 ResNet50 模型结构上的训练损失和验证精度上较为明显的改进，使得 GC 技术能够在判别器的训练过程和泛化能力上得到增强。梯度中心化算法对比在 WGAN、LS-GAN 和 GLS-GAN 等模型上对损失函数和评价图像数据的 FID 指标的变化表明：梯度中心化算法提高了 GAN 模型中生成器拟合真是样本数据生成的生成样本的图像质量和图像多样性。

4.2 Game Theory（博弈论角度）角度下的泛化能力和纳什均衡

GAN 模型训练的方式就是对抗训练，对抗在博弈论中研究比较深入。既然博弈是 GAN 模型训练动力来源的本质上的描述的话，研究 GAN 训练就可以从博弈的角度来阐述。博弈与信息是不可或缺的。G 和 D 在研究者的眼中常常被看作造假者和鉴假者。造假者要想使得赝品更加真实，他会不断地需要从鉴假者中获取信息来改进自己的赝品的质量。这其中所说的信息就是从鉴别器传入生成器的梯度信息，而梯度下降的快与慢就取决于判别器自己对数据样本的分类的泛化能力。由于博弈论中的纳什均衡比较理想化，实际经济研究中也发现这种均衡几乎从未达到过。那么梯度下降曲线是非凸函数的话，是否可以找到一个预期中的局部最优鞍点，以此达到训练 GAN 模型的目的。当然，现存的许多文献中都尝试着通过动力学、势场、迭代训练的约束和时标更新规则等等去寻找和证明那个局部最优点，仿佛数学上的证明很难确定那个点。而本文则是通过直接作用于梯度向量的方式提高判别器的泛化能力来向生成器输入尽量优的梯度信息。

5 算法结构和研究结果

5.1 算法结构

首先，生成对抗网络有两大结构：生成器（G）和判别器（D）。如图 5-1 所示：

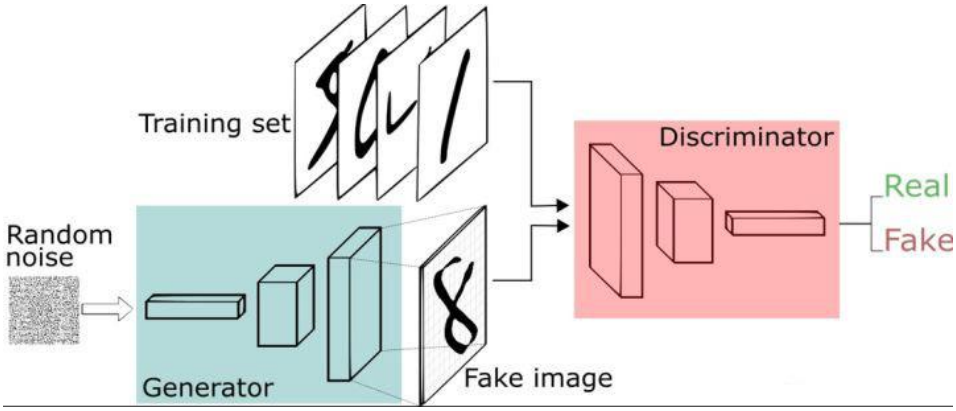


图 5-1　基本 GAN 模型结构图

上述为 GAN 模型中通用架构。其中本文中 WGAN 以及使用梯度中心化算法改进 GAN 训练速度技术也是这个结构。

在图像数据领域，当今比较成功的神经网络是卷积神经网络。其结构如图 5-2 所示。

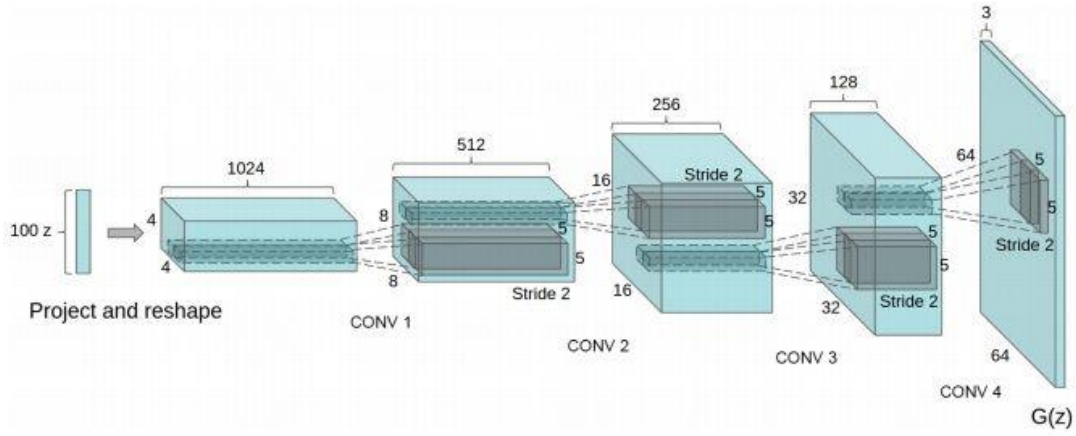


图 5-2　生成器 G 的反卷积神经网络图

GAN 模型的生成器和判别器的内部结构都是神经网络结构，只是它们中生成器是对于多维向量进行反卷积生成图像，而判别器是对于多维张量数据（如图像）进行分类判别出输入样本数据是真实的还是生成的。

GAN 模型训练的本质过程就是 G 对符合高斯分布或者其他分布的噪声经过神经网络（例如深度卷积神经网络）进行生成图像数据（生成样本）。D 经过一次或 K 次训练过后对生成样本和真实样本进行判别真假。之后，G 不断拟合真实样

本为了迷惑 D，让 D 判别不出来哪个是真和假。这种情况常见于造假者和鉴假者之间的博弈。

虽然结构只有 G 和 D，但是模型要训练起来，必须有目标函数（Loss）、噪声数据、真实数据集和优化器进行训练。超参数设置有学习率和 epoch（循环次数）。

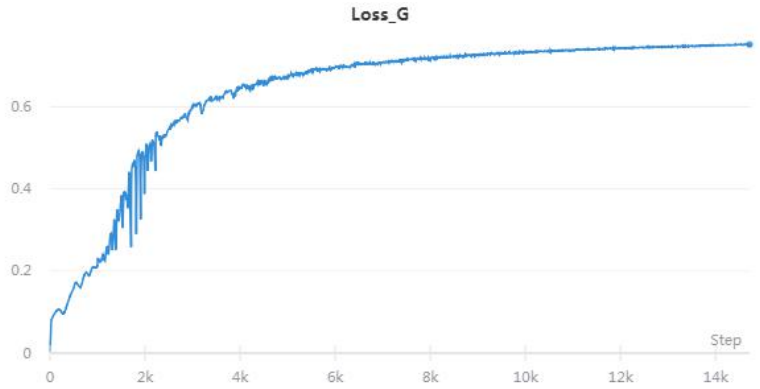
GAN 模型中的 G 和 D 两个结构组成了对抗训练的两者。其中 G 的内部结构则是对噪声向量经过反卷积（卷积网络倒回来）生成高维图像数据，最常用到的采样方式则是上采样。然而 D 的内部结构则是对高维图像数据经过多层嵌套结构将低维数据表示逐步生成高层次、抽象的特征层，继而对其进行分类是否是真实和生成样本数据。

G 和 D 内部的结构常常为神经网络，神经网络经常使用卷积神经网络及其多个变形作为图像分类应用，在文本数据的生成中常常使用循环神经网络和 LSTM 结构等等，常常在神经网络对于真实音频数据和生成的数据分布进行识别之初需要在数据预处理方面使用高斯混合模型（GMM）和隐马尔可夫模型（HMM）、其他一些的音频预处理等等进行语音识别。

5.2 WGAN-与标准 GAN 模型结构不同的算法结构

5.2.1 WGAN 不同于标准的 GAN 模型

具体对比使用基于动量优化算法和梯度中心化算法的 Adam_GCC 优化器和 SGD_GCC 优化器的生成器和判别器的损失函数下降。如图 5-3 和图 5-4 所示。



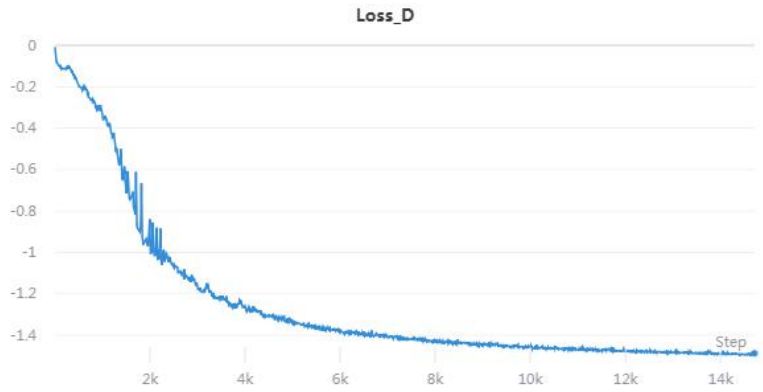


图 5-3　SGD_GCC 优化器训练过程 G 和 D 的损失函数图

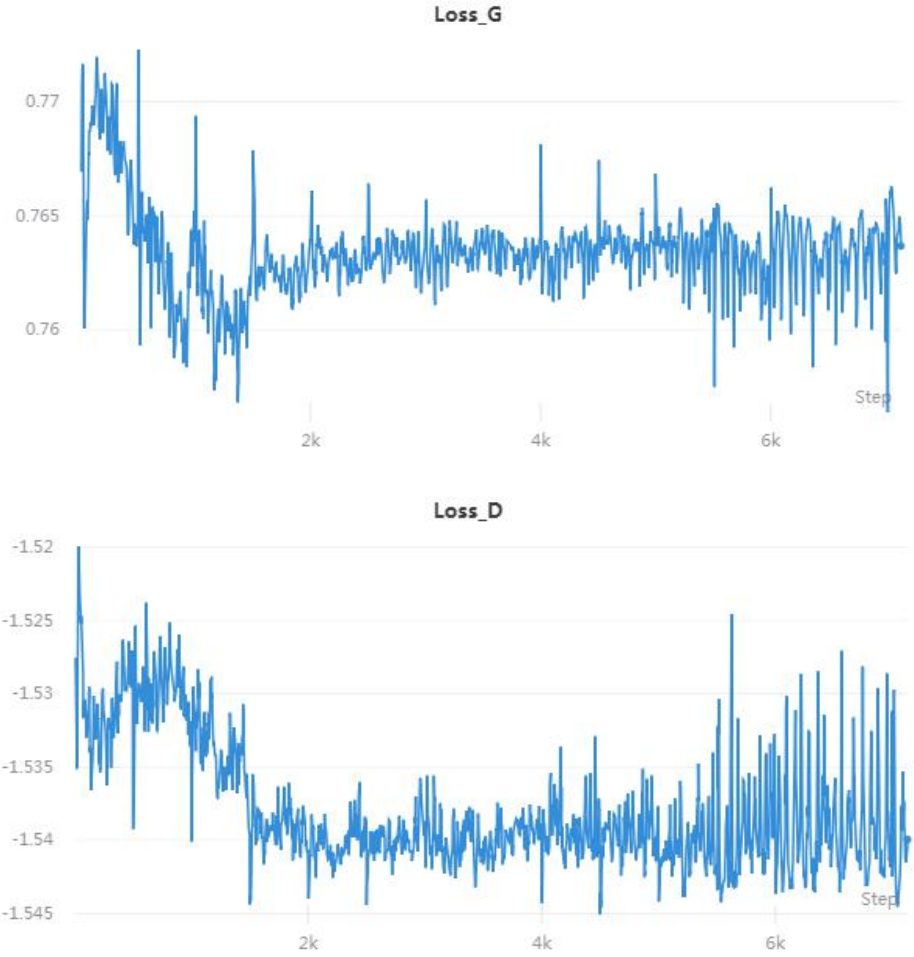


图 5-4　Adam_GCC 优化器训练过程 G 和 D 的损失函数图

由上诉图 5-3 和图 5-4 所示，在 SGD_GCC 优化器下 WGAN 模型训练过程中 G 和 D 的损失函数变化比较平滑，不易突变，由此可以表示纳什均衡状态不易发生震荡现象。而在 Adam_GCC 优化器下 G 和 D 的损失函数变化整体一致，但是突变现象比较明显，由此可表示纳什均衡状态易发生震荡现象，不易提高模型训练速度。

同时，本文又对比了带有梯度中心化算法改进的 SGD_GCC 优化器技术和 SGD 优化器技术训练之后的生成器和判别器的损失函数图，知晓了经过梯度中心化算法改进的 WGAN 模型训练使得损失函数变化更加迅速和平滑，极大地提高了模型训练速度和模型收敛，同时纳什均衡状态的震荡现象也较少。通过仔细地比较训练前期 SGD_GCC 优化技术下的损失函数变化较大也比较平滑，而在 SGD 优化技术下损失函数变化虽也平滑但变化较小，训练速度不够。具体曲线变化图如图 5-5 所示。

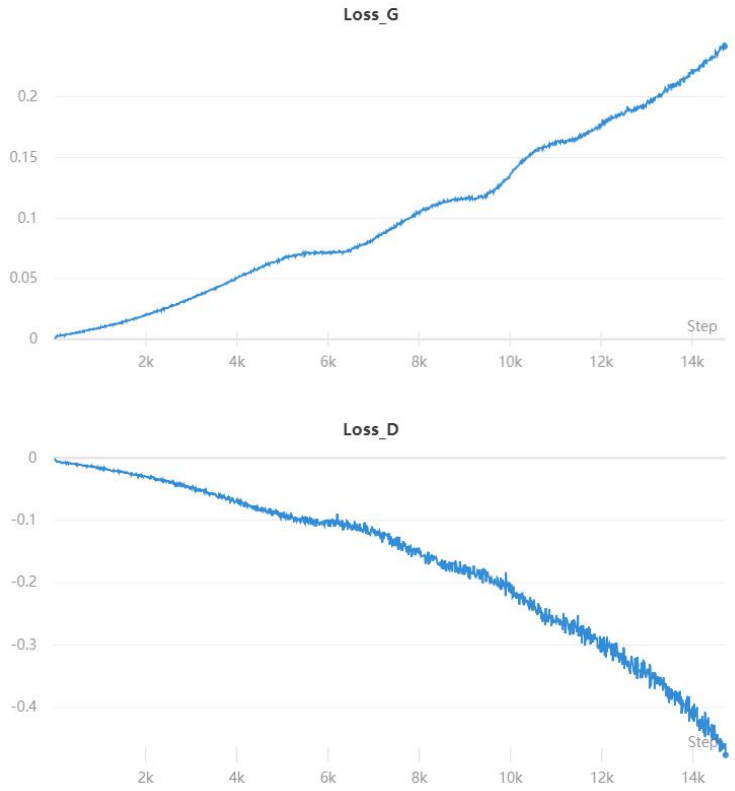


图 5-5　SGD 优化器训练过程 G 和 D 的损失函数图

6　结论

6.1　总结

自从 GAN 模型被 Ian Goodfellow 提出之后，该模型以其独特的训练方式很快地吸引了大批学者争相去研究和应用，希望以此新颖的方式突破深度学习领域的一些实际应用中的瓶颈。但是理想很美满，现实却是很骨感。在实际的应用中，

学者们发现对抗训练的方式虽然很独特且与人的认知很相似，但是这种训练方式却很难平稳地进行到底。经常出现的模式崩溃问题消耗了大量的时间却解决。经过近几年对 GAN 模型的研究与应用，出现的 Deepfake（人工智能换脸技术）和生成对抗网络与一阶动画模型相结合的“视频会议阿凡达”将会再一次引发人们对于 GAN 的关注和担忧。虽然，GAN 模型以其独特的训练方式使得计算机仿佛具有了创造能力，但是它的判别器与生成器对抗训练要达到的纳什均衡却很薄弱，且无法达到理想的那种纳什均衡。

本文内容阐述了从前人研究的目标函数的角度来加速 GAN 模型的稳定训练能力和提高图像质量和多样性的典型的 GAN 模型（WGAN、LS-GAN 和 GLS-GAN）。数学理论和实践结果相结合使得 GAN 模型的稳定训练到收敛成为了现实。上诉几种模型在 GAN 模型改进和优化中取得了万众瞩目的效果。之后本文通过一个最新的深度学习领域的训练优化器技术（梯度中心化算法）的研究和实践应用。梯度中心化算法以其直接作用在梯度向量的方法使得 GAN 模型中判别器的泛化能力得到了较大的提高，从而使得整个 GAN 模型的泛化和平衡训练得到了加强。

再者，本文通过另一种作用于梯度向量的方法（零梯度惩罚方法）和梯度中心化算法的对比研究，实践结果更加证实了提高 GAN 模型稳定训练收敛的能力可以凑够梯度向量的角度进行研究和改进。

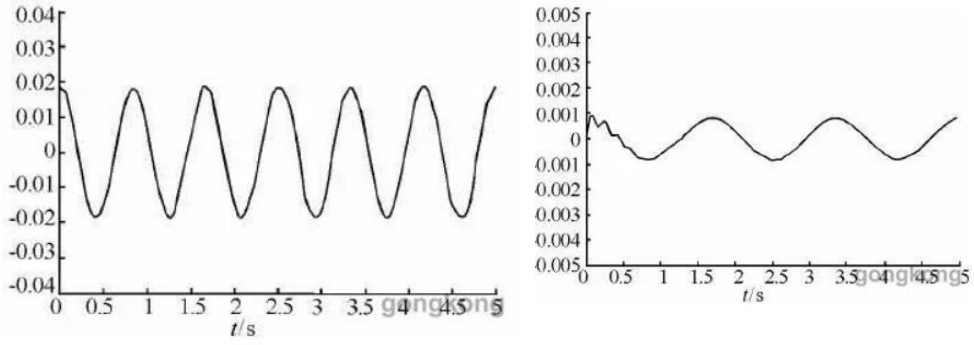
最后，通过 GAN 模型中生成器和判别器之间的对抗训练本质上是极大极小零和博弈的思想，从而尝试从博弈论的角度来解释 GAN 模型泛化能力和纳什均衡之间的联系。

6.2 研究价值和实际应用的展望

此次研究，通过两种不同的角度（目标函数和梯度向量角度）对 GAN 模型训练的理论性研究，使得应用创新角度的 GAN 模型（例如 CGAN、StytleGAN 和 InfoGAN 等等）更加专注于实际应用场景的开发研究。从而，只要应用理论上的成果使得应用模型训练更加快速和减少 GPU 计算资源的消耗，加速了产品从研发到落地部署的进程。

6.3 研究思想后续和可行性探究

研究表明将模糊神经网络与机器人控制结合在一起，达到了能够很好地控制机器人轨迹的效果。由于神经网络无需人工定义规则和能适应系统复杂多变的动态特性使得其于机器人控制有着很好的结合。这个世界的规则是复杂且有序的但又模棱两可的，那么无论现有的进行标注过的数据再怎么多都无法完全地描绘这个世界的规则。因此计算机需要从数据中自适应学习，而且需要模糊理论来试着去拟合这个复杂的世界。在《模糊神经网络在机器人控制中的应用研究》一文中，学者应用模糊理论的效果也如下图所示取得了显著的效果[11]。其如图 6-1 所示：



(a) PID 控制　　(b) 模糊神经网络控制

图 6-1　常规 PID 控制和模糊神经网络控制对比图

上诉两图分别是常规的 PID 控制和模糊神经网络控制。由上图分析可得，模糊神经网络控制的振荡幅度较小。因此遐想，常见的 GAN 模型训练过程所要达到的纳什均衡普遍存在振荡现象。那么是否可以在 GAN 模型的生成器和判别器内部神经网络结构加入模糊理论结构？是否能够达到预期中减小纳什均衡振荡现象的振荡幅度以此来提高 GAN 模型训练速度？由于现存的文献参考中对模糊理论研究很少，再加上理论上纳什均衡振荡现象无法描述和模糊神经网络的实际应用较少使得本文对其技术不再赘述。不过，在直觉模糊二人非合作博弈理论中对于模糊理论与博弈论的理论上描述详尽或许在之后的研究中起到理论上对于纳什均衡振荡现象的研究起到不小的帮助。

参考文献:

[1] Oliehoek, Frans A., et al. "Beyond local nash equilibria for adversarial

networks." Benelux Conference on Artificial Intelligence. Springer, Cham, 2018.
[2] Goodfellow, Ian, et al. "Generative adversarial nets." Advances in neural information processing systems. 2014.
[3] Qi, Guo-Jun. "Loss-sensitive generative adversarial networks on lipschitz densities." International Journal of Computer Vision (2019): 1-23.
[4] Arora, Sanjeev, et al. "Generalization and equilibrium in generative adversarial nets (gans)." Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org, 2017.
[5] Miyato, Takeru, et al. "Spectral normalization for generative adversarial networks." arXiv preprint arXiv:1802.05957 (2018).
[6] Chen, Ting, et al. "Self-supervised gans via auxiliary rotation loss." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.
[7] Yong, Hongwei, et al. "Gradient Centralization: A New Optimization Technique for Deep Neural Networks." arXiv preprint arXiv:2004.01461 (2020).
[8] Thanh-Tung, Hoang, Truyen Tran, and Svetha Venkatesh. "Improving generalization and stability of generative adversarial networks." arXiv preprint arXiv:1902.03984 (2019).
[9] Heusel, Martin, et al. "Gans trained by a two time-scale update rule converge to a local nash equilibrium." Advances in neural information processing systems. 2017.
[10] Arjovsky, Martin, Soumith Chintala, and Léon Bottou. "Wasserstein gan." arXiv preprint arXiv:1701.07875 (2017)