2018-12-25，**主要内容：**

深度学习初稿撰写要求：

1. 不能直接粘贴拷贝网上或者其他书籍内容，需按照自己的理解撰写。
2. 图表不能用现有图表，需重新制作，图用Vision绘图，表用Word制表。
3. 每章后需整理本章参考文献
4. 不局限于提纲内容，有修改或者新加内容随时和我讨论
5. 图文并茂，尽量通俗易懂 。

# 第七章 其他典型深度学习方法

**7.1 生成对抗网络**

## 7.1.1概述

生成对抗式网络（Generative Adversarial Networks，GANs）是一种以无监督方式学习目标分布的深度生成模型。这种目标分布的学习能广泛的应用于很多应用，包括：文字转化图形、图像生成、图像风格转换、图像重构与分类等。GANs最早在2014年提出，是一种半监督与无监督的学习方法，它实现通过隐性数据的高维分布。GANs是训练一对相互竞争的网络，基于博弈论的思想。这里我们做一个形象地比喻，可以将其中一个网络视为造假者（生成器），另一个看作专家。学习的过程就是造假者尽最大努力仿造出接近真实的物品，专家（判别器）的职责就是将仿造品与真品进行比较，鉴别真伪。重复以上过程，专家和伪造者同时训练并相互竞争，直到两者达到一个均衡和谐的状态后，最终伪造者能产出更为真实的的物品，专家则一个判断能力较强的分类器。前者可以用于机器创作（自动画出“猫”“狗”），而后者则可以用来机器分类（自动判断“猫”“狗”）。

## 7.1.2基本算法原理

7.1.2.1 基本结构

GANs基于博弈论的思想，其中生成器网络（generator network）与其对手判别器网络（discriminator network）相互竞争。 生成对抗网络是一种生成式的建模方法，GANs的结构和我们之前见到的神经网络略为不同。大体上来说，GANs由生成模型Generator和判别模型Discriminator组成，基本的结构图如下：



其中，生成模型我们可以将其看做一个神经网络模型，输入一个噪声/样本后，输出的将是一个图像。上面结构图中包括两个数据集一个真实数据集另一个合成数据集，即生成网络合成的猫图像数据。判别模型，也是一个简单的神经网络架构，输入是一副图像，分别来自真实训练数据和合成数据，通过判别网络后输出则是一个概率值，基于此来判断猫图像的真假程度。随着以上训练的进行，判别器‘被迫’增强自身的判别能力，而生成器‘被迫’生成越来越逼真的输出，以欺骗判别器。理论上，最终生成器和判别器会达到一种均衡“纳什均衡”，即生成网络合成的假样本数据进入判别网络后，判别网络输出的结果是近似接近0.5的值，不能区分真假样本。

7.1.2.2 形式化表示

GANs中生成器网络G采用以随机噪声/样本为输入的多层神经网络，判别器网络D也是一个多层神经网络。G可以表示成 ,其中z为随机噪声，服从分布。D输入一是个真实图像及合成（虚假）图像x, 输出结果可以表示为表示输入图像是真实和伪造图像。根据训练过程我们可以定义一个损失函数：

**（公式7.1）**

其中x、z分别表示来自训练集数据和随机噪声数据，优化目标是。

为便于理解，我们对上面公式做一些简单说明，是指让D的损失越大越好, 越大说明D越能发现G生成的数据是假的, 因为二者目的就是给对方制造麻烦的. 是指G希望D的损失越小越好, 因为越小说明G生成的数据越能骗过D。

是希望**(公式7.1)**最大, 当为1,为0时,说明D把所有真实的数据识别出来了,并把所有的假的数据也识别出来, 此时，我们便得到一个优秀的判别器，损失达到最大。相反，是希望**(公式7.1)**最小, 当时,判别器就把所有的假数据当作了真的, 此时，说明G非常成功，生成器就就具备了以假乱真的功能。

  总之，我们在更新D和G的参数时就是分别对D进行梯度上升更新,对G进行梯度下降更新。 到最后达到一个动态平衡, 也就是D的输出为, 即D已经不能区分真假了。

7.1.2.3 算法的优势和不足（https://blog.csdn.net/xg123321123/article/details/78034859）

1. 优势
2. 模型只用到了反向传播,而不需要马尔科夫链。
3. 理论上,只要是可微分函数都可以用于构建D和G,因为能够与深度神经网络(CNN、RNN等)结合做深度生成式模型。
4. G的参数更新不是直接来自数据样本,而是来自D的反向传播。
5. 训练时不需要对隐含变量做推断。
6. GANs是一种以半监督方式训练分类器的方法.在你没有很多带标签的训练集的时候,你可以不做任何修改的直接使用我们的代码,通常这是因为你没有太多标记样本。
7. 相比玻尔兹曼机,GANs生成实例的过程只需要模型运行一次,而不是以马尔科夫链的形式迭代很多次。
8. 不足
9. 可解释性差,生成模型的分布Pg(G)没有显式的表达。
10. 比较难训练,D与G之间需要很好的同步，例如D更新k次而G更新一次。GAN的学习过程可能发生崩溃问题，生成器开始退化，总是生成同样的样本点，无法继续学习。当生成模型崩溃时，判别模型也会对相似的样本点指向相似的方向，训练无法继续。
11. 网络难以收敛，目前所有的理论都认为GAN应该在纳什均衡上有很好的表现，但梯度下降只有在凸函数的情况下才能保证实现纳什均衡。
12. 对于生成离散数据比较困难,就像文本。

## 7.1.3应用（https://www.jianshu.com/p/80bd4d4c2992）

GANs是一个生成模型，其最直接的应用就是真实数据的建模与生成，比如音频、图像等。同时，可在一定程度上解决传统机器学习训练数据不足的问题。总之，目前无论在学术界和还是在工业界，如推特、苹果公司等，GANs都有广泛的应用。

1. 文本到图像

实通过输入目标图像的相关属性的描述性文本后，GANs可输出与之对应的最为真的图片。

1. 图像超分辨率

图像超分辨率一直是一个很重要的研究课题，比较重要的是对天文图像和卫星图像做超分辨率，不管是在天文，军事还是其他方面，都有很重要的应用。在生活中，如果有标清的视频可以变为高清的视频

1. 图像翻译

指的是图像到图像的翻译，比如说将语义标注图、灰度图或边缘图作为GAN的输入，那么我们希望它输出能够和输入图一致的真实图像，例如这里的街景图和彩色图。

**参考文献**

⑴ Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio.2014. Generative Adversarial Networks.

(2) Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, Anil A Bharath.2017. Generative Adversarial Networks: An Overview.

7.6 深度森林

7.6.1概述

7.6.2 主流算法原理

7.6.3应用场景

**参考文献**

# 第十一章 深度学习在语音识别中的典型应用

11.1 声学模型

**11.1.1 概述**

**语音识别系统主要有四部分组成：信号处理、特征提取、声学模型、解码器**

**传统的声学模型：HMM-GMM 、CD-DNN-HMM**

**基于神经网络的模型：DNN-HMM、CTC算法（Connectionist temporal classification）**

**11.1.2主流算法**

**传统**

**深度学习**

**11.1.3场景应用**

**11.1.4 训练自己的声学模型：**

**需有数据集和代码**

**参考文献**