

DOI:10.13364/j.issn.1672-6510.20210070

数字出版日期: 2021-09-29; 数字出版网址: <http://kns.cnki.net/kcms/detail/12.1355.N.20210928.1129.005.html>

基于对抗性策略的变分自编码模型

陈亚瑞, 王浩楠, 张芝慧, 杨剑宁, 丁文强

(天津科技大学人工智能学院, 天津 300457)

摘要: 深度生成模型结合了生成模型与多层神经网络, 在计算机视觉、目标检测、自然语言处理等领域得到广泛应用. 变分自编码是一种重要的深度生成模型, 通过多层神经网络对生成模型和后验概率分布的近似分布进行建模, 得到目标函数的变分下界. 变分自编码以其显式的生成模型建模方式及定量的下界表示形式受到研究者的关注. 由于变分自编码中推理模型的表示能力有限, 使得生成的图像精细度不高. 本文提出一种基于对抗性策略的变分自编码模型, 该模型通过在变分自编码模型的推理模型中加入随机噪声, 提高推理模型的表示能力, 同时引入对抗网络增加聚合正则化约束, 进一步训练推理模型. 通过在 MNIST 和 Fashion-MNIST 数据集上设计对比实验, 表明该模型可以给出更优的变分似然下界, 并生成效果更好的图像数据.

关键词: 生成模型; 变分自编码; 对抗策略; 变分似然下界; 图像生成

中图分类号: TP391

文献标志码: A

文章编号: 1672-6510(2022)01-0056-08

Variational Autoencoder Model Based on Adversarial Strategy

CHEN Yarui, WANG Haonan, ZHANG Zhihui, YANG Jianning, DING Wenqiang

(College of Artificial Intelligence, Tianjin University of Science & Technology, Tianjin 300457, China)

Abstract: Deep generative models combine generative models and multilayer neural networks, and they are widely used in computer vision, target detection and data dimensionality reduction. Variational autoencoder is an important deep generative model. Modeling the generative model and the variational approximate distribution of the posterior probability distribution through a multilayer neural network can obtain the lower bound of the variation of the objective function. Variational autoencoder has attracted the attention of researchers with its explicit generative model modeling and quantitative lower bound representation. However, due to the limited representation ability of the inference model in the variational autoencoder, the fineness of the generated image is not high. Therefore, in this paper we propose a variational autoencoder model based on adversarial strategy. The model adds random noise to the inference model of the variational autoencoder to improve the representation ability of the inference model. At the same time, the adversarial network is introduced to increase the aggregation regularization constraint for further training inference model. The experimental results on MNIST and Fashion-MNIST datasets show that the model can obtain better variational lower bound of the variational likelihood under given parameters, and the effect of image generation is also improved.

Key words: generative model; variational autoencoder; adversarial strategy; variational lower bound of the variational likelihood; image generation

生成模型(generative model)是通过引入隐变量建模数据的生成过程, 训练后的模型可以用于生成新的数据样本. 深度生成模型是生成模型和深度神经网络相结合. 变分自编码(variational autoencoder,

VAE)模型^[1-2]和生成对抗网络(generative adversarial networks, GAN)^[3-4]作为应用最广泛的两类深度生成模型, 近年来凭借强大的性能表现受到图像处理^[5]、计算机视觉、自然语言处理等^[6]领域的关注. 这两类

收稿日期: 2021-04-01; 修回日期: 2021-05-27

基金项目: 天津科技大学青年教师资助计划项目(2017LG10); 天津市研究生科研创新项目(2020YJSZXS31)

作者简介: 陈亚瑞(1982—), 女, 河北邢台人, 副教授, yrchen@tust.edu.cn

深度生成模型能对复杂的高维度数据进行建模, 而且随着它们变体模型的产生^[7-8], 已经在数据特征提取、数据降维、目标检测和生成新数据集等方面取得较好的实际效果. 变分自编码模型通过对生成过程进行显式建模, 采用似然函数为优化目标, 其中基于变分推理的证据下界优化度量指标为衡量模型优劣提供了定量指标, 该模型已经成为深度生成模型领域的研究热点.

变分自编码器模型作为无监督特征学习^[9]的重要工具, 其主要思想是通过神经网络对生成模型和后验概率分布的变分近似分布进行建模, 通过学习隐空间与数据空间的特征映射, 输出新的重构数据空间. 但因其推理模型表示能力不足, 所以生成图像较为模糊, 生成样本多样性受限. 当前很多研究工作基于变分自编码模型进行改进, 提升数据生成质量及生成数据的多样性. Manuel 等^[10]将类别标签信息加到变分自编码模型中, 提出条件变分自编码 (condition variational autoencoder, CVAE) 模型, 并将其应用于半监督学习中. Louizos 等^[11]为了让变分自编码模型能够更明确地学习到某些非变性因素的特征表示, 提出变分公平自编码 (variational fair autoencoder, VFAE) 模型. Vincent 等^[12]为了去除生成图像的噪点, 改善变分自编码的生成效果, 提出了去噪变分自编码 (denoising variational autoencoder, DVAE) 模型. Kingma 等^[13]将自回归神经网络 (如 PixelRNN^[14]) 应用于变分自编码模型中, 充分利用自回归神经网络的生成能力, 极大地提升了变分自编码模型的生成效果.

生成对抗网络^[3-4]的提出受到业界的广泛关注. 因其采用对抗训练的方式, 能够生成更清晰、更真实的数据样本, 进而被应用于无监督和半监督学习领域^[4,15]. 然而, 生成对抗网络中的训练不稳定、梯度消失和模式崩溃等问题有待解决. 为了优化变分自编码模型和生成对抗网络, 有许多研究人员将生成对抗网络的对抗策略引入到变分自编码模型中. Makhzani 等^[7]将对抗策略引入到自编码模型中, 提出对抗自编码 (adversarial autoencoder, AAE) 模型, 其度量模型精度的方式已不再是变分下界, 虽然没有广泛沿用, 但模型思想具有很高的研究价值. Larsen 等^[16]将生成对抗网络引入到自编码模型中, 提出一种优化像素级图像的混合模型 (AE-GAN). Dumoulin 等^[17]提出对抗学习推断理论 (adversarially learned inference), 将对抗策略引入到推理模型中, 让对抗策

略与变分自编码模型的结合进一步成为可能. Mescheder 等^[8]将对抗思想引入到变分贝叶斯模型中, 用“黑盒方法”重新定义推理模型, 提出对抗变分贝叶斯 (adversarial variational Bayes, AVB) 模型. Huang 等^[18]提出 Intro 变分自编码 (introspective variational autoencoders, IntroVAE) 模型, 并通过该模型得到了超高分辨率的图像生成效果.

在变分自编码模型及其变体模型中, 推理模型的代表能力不足一直是生成图像效果不佳的重要原因. 因此, 本文提出基于对抗策略的变分自编码 (adversarial variational autoencoder, AVAE) 模型解决这一问题. 首先介绍变分自编码模型和对抗变分贝叶斯模型, 其次对提出的基于对抗策略的变分自编码模型进行详细介绍, 最后结合实验验证该生成模型在给定参数的情况下能够增加推理模型的代表能力, 得到更好的变分似然下界, 有效提升图像的生成效果.

1 背景

1.1 变分自编码模型

传统变分自编码模型是由推理模型 (又称编码器) 和生成模型 (又称解码器) 组成的^[1-2], 其中: 推理模型是通过多层神经网络将真实数据 x 编码为一个低维隐变量 z , 生成模型是将隐变量 z 通过多层神经网络还原映射到高维度数据空间. 变分自编码模型结构如图 1 所示, 其中白色节点 z 表示隐变量, 灰色节点 x 表示可观测量, 节点之间的有向线段表示变量之间的依赖关系; 黑色实心小方块表示模型参数值; $q_{\phi}(z|x)$ 与虚线部分为推理过程, $p_{\theta}(z|x)$ 与实线部分为生成过程, ϕ 和 θ 为相关过程的参数; 方框表示该过程可以重复出现, 例如在该模型中基于隐变量可以重复生成数据样本, 观测到 N 条数据, 则该过程重复出现了 N 次.

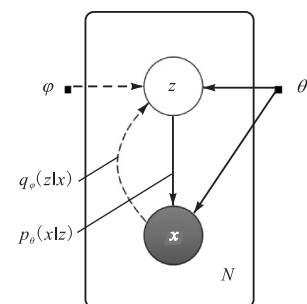


图 1 变分自编码模型

Fig. 1 Variational autoencoder model

在变分自编码模型中,先验概率分布 $p(z)$ 服从标准高斯分布 $N(0, I)$, 推理模型通过引入近似后验概率分布 $q_\varphi(z|x)$ 逼近真实后验概率分布 $p_\theta(z|x)$, φ 和 θ 均为模型参数. 通过推导, 得到数据对数似然 $\log p_\theta(x)$ 的下界, 又称为证据下界^[2]或变分似然下界, 即

$$\log p_\theta(x) \geq L(\theta, \varphi; x) = E_{q_\varphi(z|x)} \log p_\theta(x|z) - D_{\text{KL}}(q_\varphi(z|x) \| p(z)) \quad (1)$$

其中: $E_{q_\varphi(z|x)} \log p_\theta(x|z)$ 为生成模型的重构误差项, 用来比较新生成的图像与真实图像之间的差距, 通过最大化重构误差项, 模型可以得到更好的生成效果; $D_{\text{KL}}(q_\varphi(z|x) \| p(z))$ 为两个分布之间的 KL 散度, 用 D_{KL} 进行正则化约束, 通过让近似后验概率分布逼近给定的先验分布, 使正则化项最小.

此时, 变分自编码模型的目标是通过随机梯度下降算法^[19]学习到最优的模型参数 φ 和 θ , 使得证据下界最大, 即

$$\theta, \varphi = \arg \max_{\theta, \varphi} L(\theta, \varphi) \quad (2)$$

变分自编码模型在训练过程中使用了重参数化方法 (reparameterization)^[1-2], 目的是解决随机梯度下降过程中参数的梯度消失问题.

作为生成模型, 虽然变分自编码模型将变分下界作为模型精度度量标准, 但是在图像生成过程中仍存在生成图像模糊和生成样本多样性受限等问题. 这些问题与近似推理模型的表示能力有限有关^[16, 18], 而对抗变分贝叶斯模型很大程度上解决了这一问题.

1.2 对抗变分贝叶斯模型

对抗变分贝叶斯模型^[8]通过引入“黑盒方法”增加推理模型的表示能力, 然后引入对抗策略求解目标函数中的正则化项. 该模型主要包含 3 个组件, 分别是推理模型、生成模型及鉴别器, 如图 2 所示: Q_φ 为推理模型, 将真实图像 x 通过神经网络编码为隐变量 z ; P_θ 为生成模型, 将隐变量 z 还原映射到高维度数据空间; D 为鉴别器, 用于鉴别隐变量与图像数据形成的数据组合; φ 和 θ 为相关组件的学习参数.

对抗变分贝叶斯模型将先验概率分布 $p(z)$ 设定为标准高斯分布 $N(0, I)$, 将真实图像 x 产生的隐变量 z 输入到生成模型中, 得到重构图像 x' 并计算负重构误差. 在此期间, 对抗变分贝叶斯模型重新定义优化为

$$\max_{\theta, \varphi} E_{p_{\text{Data}}(x)} E_{q_\varphi(z|x)} [\log p(z) - \log q_\varphi(z|x) + \log p_\theta(x|z)] \quad (3)$$

并通过“黑盒方法”向推理模型中加入随机噪声, 增加推理模型的灵活性, 能有效地表示任何种类隐变量的分布, 进而将式 (3) 中 $\log p(z) - \log q_\varphi(z|x)$ 与鉴别器建立联系, 将其推导成含有鉴别器函数 $D(x, z)$ 的形式, 得出优化参数 φ 时的最优解为

$$D^*(x, z) = \log q_\varphi(z|x) - \log p(z) \quad (4)$$

至此, 对抗变分贝叶斯模型的目标函数为

$$\max_{\varphi, \theta} E_{p_{\text{Data}}(x)} E_{q_\varphi(z|x)} (\log p_\theta(x|z) - D^*(x, z)) \quad (5)$$

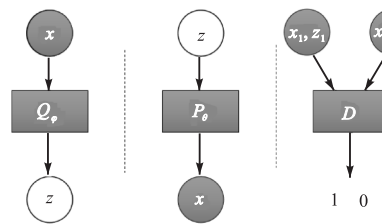


图 2 对抗变分贝叶斯模型结构示意图

Fig. 2 Schematic diagram of the structure of adversarial variational Bayes model

模型的鉴别器有两个输入, 分别是真实图像与隐变量组合 (x_1, z_1) 和真实图像与先验随机变量组合 (x_2, z_2) . 变分贝叶斯模型认定 (x_1, z_1) 为正样本, (x_2, z_2) 为负样本, 以此让模型引入对抗的训练方式. 鉴别器的目标函数为

$$\max_D E_{p_{\text{Data}}(x)} E_{q_\varphi(z|x)} \log(D(x, z)) + E_{p_{\text{Data}}(x)} E_{p(z)} \log(1 - D(x, z)) \quad (6)$$

对抗变分贝叶斯模型在训练过程中让推理模型和鉴别器交替训练, 持续提升推理模型的表示能力和生成模型的生成能力. 待模型训练完成后, 可以将先验分布采样的随机噪声通过生成模型直接映射到数据空间, 以此作为模型的最终输出. 该模型生成图像的效果相比于变分自编码模型有很大提升, 且随着训练的进行, 生成的图像也更具多样性.

2 改进的变分自编码模型

2.1 引入噪声的推理模型

在传统变分自编码模型的推理模型中引入随机噪声, 将推理模型表示成 $q_\varphi(z|x, \epsilon_1)$, 其中 ϵ_1 是来自标准高斯分布 $N(0, I)$ 的随机噪声, φ 为推理模型参数. 具体做法是将低维度的随机噪声加到高维度的

图像数据中, 增加推理模型输入端的维度. 此时推理模型中不仅包含图像数据, 而且包含低维的随机噪声, 可提升其表示能力. VAE 和 AVAE 推理模型示意图如图 3 所示, 其中 Q_φ 为推理模型, x 为图像数据, z 为隐变量, ε_1 为来自标准高斯分布的随机噪声, 即推理模型噪声. 模型的优化目标为

$$L(\theta, \varphi) = E_{p_{\text{Data}}} \left[E_{q_\varphi(z|x, \varepsilon_1)} p_\theta(x|z) - D_{\text{KL}}(q_\varphi(z|x, \varepsilon_1) \| p(z)) \right] \quad (7)$$

其中: φ 和 θ 分别为推理模型和生成模型的参数. AVAE 模型通过在推理模型中引入随机噪声, 使其获得学习更复杂数据分布的能力, 进而提升推理模型的代表能力.

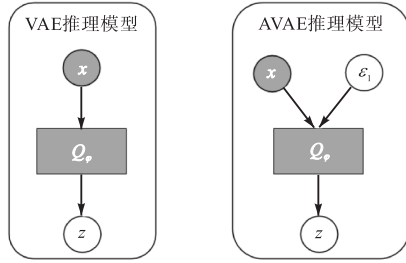


图 3 VAE 和 AVAE 推理模型示意图

Fig. 3 Schematic diagram of VAE and AVAE inference model

AVAE 模型沿用变分自编码模型中的重参数化方法^[2,11]避免模型在训练过程中遇到参数的梯度消失, 进而能更顺利地使用随机梯度下降对参数更新. 模型重参数化过程如图 4 所示, 其中 x 为图像数据, z 为隐变量, μ 和 σ 为推理模型的神经网络计算出的隐变量分布的均值和方差, ε_2 为来自标准高斯分布的重参数化噪声, g_φ 为重参数化函数, 其形式为 $g_\varphi(\varepsilon, x) = \mu + \sigma \odot \varepsilon$.

重参数化方法是对于每一个隐变量 $z \sim q_\varphi(z|x, \varepsilon_1)$, 通过使用随机变量 ε_2 的可微分函数 $g_\varphi(\varepsilon_2, x)$ 重新设定隐变量 z

$$z = g_\varphi(\varepsilon_2, x), \quad \varepsilon_2 \sim N(0, I) \quad (8)$$

对于每一个真实数据 x , 采用随机梯度下降算法, 将式(7)改写为

$$\tilde{L}(\varphi, \theta) = \frac{1}{L} \sum_{l=1}^L \log p_\theta(x|z^{(l)}) - D_{\text{KL}}(q_\varphi(z|x, \varepsilon_1) \| p(z)) \quad (9)$$

其中: $-D_{\text{KL}}(q_\varphi(z|x, \varepsilon_1) \| p(z))$ 为模型的正则化项, $\frac{1}{L} \sum_{l=1}^L \log p_\theta(x|z^{(l)})$ 为负重构误差项. 通过使用

$z = g_\varphi(\varepsilon_2, x)$ 函数, 将真实数据 x 和随机噪声 $\varepsilon_2^{(l)}$ 映射到该数据的近似后验样本.

$$z^{(l)} = g_\varphi(\varepsilon_2^{(l)}, x) = \mu + \sigma \odot \varepsilon^{(l)}, \quad \varepsilon_2^{(l)} \sim N(0, I) \quad (10)$$

将样本 $z^{(l)}$ 输入到式(9)中等号后边第二项, 即推导出生成模型生成真实数据 x 的概率密度. 基于随机噪声的模型优化问题为

$$\varphi, \theta = \arg \max_{\varphi, \theta} \tilde{L}(\varphi, \theta) \quad (11)$$

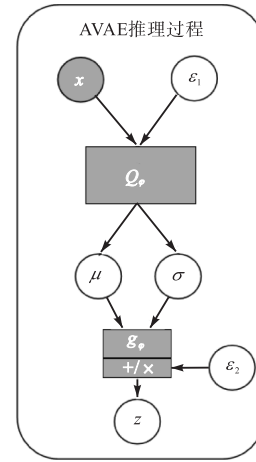


图 4 AVAE 模型的重参数化及推理过程

Fig. 4 Reparameterization and inference process of AVAE model

2.2 对抗网络

AVAE 模型通过增加对抗网络引入对抗策略, 同时增加聚合正则化约束, 进而更好地训练推理模型. 推理模型 $q_\varphi(z|x)$ 表示单个样本 x 的近似后验概率分布, 对于样本集的隐变量空间可以用聚合后验概率分布 $q(z)$ 表示, 即

$$q(z) = \int_x q_\varphi(z|x) p_{\text{Data}}(x) dx \quad (12)$$

AVAE 模型在对抗网络的输入中加入图像数据 x , 如图 5 所示, 其中 D_ϕ 为对抗网络部分, z 为隐变量, z' 为从先验分布 $p(z)$ 中采样的随机变量. 对抗网络将图像数据与先验随机变量的组合 (x, z') 判定为正样本, 将图像数据与隐变量的组合 (x, z) 判定为负样本.

由此, 对抗网络的目标函数为

$$L(\varphi, \phi) = E_{p_{\text{Data}}(x)} E_{p(z)} \log(D_\phi(x, z)) + E_{p_{\text{Data}}(x)} E_{q_\varphi(z|x, \varepsilon_1)} \log(1 - D_\phi(x, z)) \quad (13)$$

其中: $D_\phi(x, z)$ 为对抗网络函数, ϕ 表示对抗网络参

数. 对抗网络的目的是增加聚合正则化项, 让聚合后验分布 $q(z)$ 近似给定的先验分布 $p(z)$. 因为增加了聚合正则化约束, 模型可以通过对抗训练进一步优化推理模型参数 ϕ . 对抗网络在模型下的优化问题为

$$\phi, \phi = \arg \max_{\phi} \arg \min_{\phi} L(\phi, \phi) \quad (14)$$

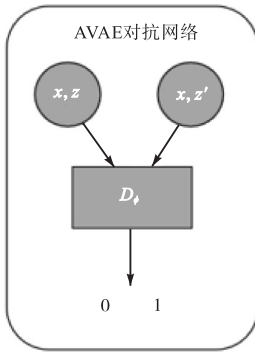


图5 AVAE模型对抗网络示意图

Fig. 5 Schematic diagram of AVAE model's adversarial network

2.3 AVAE算法

AVAE 模型通过向变分自编码模型的推理模型增加随机噪声, 增加推理模型的代表能力, 同时引入对抗网络增加聚合正则化约束进一步训练推理模型. 模型给定先验分布 $p(z)$ 和随机噪声 ϵ_1 的分布均为标准高斯分布 $N(0, I)$, 推理模型为 $q_{\phi}(z|x, \epsilon_1)$, 生成模型为 $p_{\theta}(x|z)$, 使用重参数化方法及随机梯度下降算法对参数进行更新.

模型先将图像数据 x 和随机噪声 ϵ_1 同时输入到推理模型中, 增加推理模型的输入维度, 提升其表示能力. 在训练过程中, 推理模型优化式 (11) 和对抗网络优化式 (14) 交替进行优化. 训练分两个阶段, 通过式 (11) 计算并优化推理模型的参数 ϕ , 以最大程度让聚合后验概率分布 $q(z)$ 近似给定的先验分布 $p(z)$; 通过式 (14), 先计算并优化对抗网络参数 ϕ , 让对抗网络与推理模型进行对抗训练, 接着再一次更新推理模型参数 ϕ , 完成聚合正则化约束, 增加推理模型的代表能力.

在训练过程中, 生成模型参数 θ 通过式 (11) 不断得到优化. 待模型训练完成后, 可以直接通过对 $p(z)$ 分布随机采样作为隐变量, 并借助训练好的生成模型生成更真实、更清晰的新图像数据. AVAE 模型算法如下:

1. $i \leftarrow 0$.
2. WHILE 没有收敛 DO.

3. 从真实数据分布 $p_{\text{Data}}(x)$ 采样 $\{x^{(1)}, x^{(2)}, \dots, x^{(k)}\}$.

4. 从第 1 个随机噪声分布中采样 $\{\epsilon_1^{(1)}, \epsilon_1^{(2)}, \dots, \epsilon_1^{(k)}\}$.

5. 从先验分布 $p(z)$ 中采样 $\{z^{(1)}, z^{(2)}, \dots, z^{(k)}\}$.

6. 对于优化式 (11) 计算参数的随机梯度.

$$g_{\theta} \quad g_{\phi} \leftarrow$$

$$\frac{1}{k} \sum_1^k \nabla_{\theta} \nabla_{\phi} \left[-D_{\text{KL}}(q_{\phi}(z_1^{(k)} | x^{(k)}, \epsilon_1^{(k)}) || p(z^{(k)})) + \log p_{\theta}(x^{(k)} | z_1^{(k)}) \right]$$

7. 基于随机梯度下降算法, 可以得到参数的更新式 $\theta \leftarrow \theta + h_i g_{\theta}$, $\phi \leftarrow \phi + h_i g_{\phi}$.

8. 对于优化式 (14) 计算参数的随机梯度.

$$g_{\theta} \quad g_{\phi} \leftarrow$$

$$\frac{1}{k} \sum_1^k \nabla_{\phi} \nabla_{\phi} \left[\log D_{\phi}(x^{(k)}, z^{(k)}) + \log(1 - D_{\phi}(x^{(k)}, z_1^{(k)})) \right]$$

9. 基于随机梯度下降算法, 可以得到参数的更新式 $\phi \leftarrow \phi + h_i g_{\phi}$, $\phi \leftarrow \phi + h_i g_{\phi}$.

10. $i \leftarrow i + 1$.

11. END WHILE.

3 实验

3.1 实验设计

本节所用的数据集包括手写数字数据集 MNIST(数据来源: <http://yann.lecun.com/exdb/mnist/>) 和服饰商品数据集 Fashion-MNIST(数据来源: <http://www.worldlink.com.cn/en/osdir/fashion-mnist.html>) 两种. MNIST 数据集包括手写数字图像 0~9, 共 10 类数据, 其中训练图像 60 000 张, 测试图像 10 000 张. 这些手写数字图像都经过尺寸标准化, 固定大小为 28 像素 \times 28 像素, 每个像素值为 0~1, 每个图像都被平展并转换为 784(28 \times 28) 个特征的一维 NumPy 数组. Fashion-MNIST 数据集包括 T-shirt(T 恤)、pullover(套衫)、dress(裙子)、coat(外套)、sandal(凉鞋)等服饰商品, 共 10 类数据, 其中训练图像 60 000 张, 测试图像 10 000 张. 这些服饰图像也经过尺寸标准化, 图像固定大小为 28 像素 \times 28 像素. 实验在 tensorflow 1.15 环境下进行, 操作系统 Windows 10, 运行平台工具 Python 3.7.

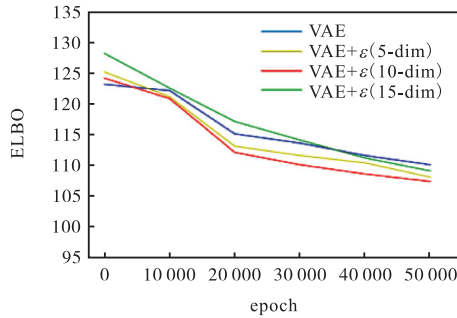
实验中 AVAE 模型的推理模型、生成模型和对抗网络均采用三隐层全连接神经网络, 中间节点个数分别为 256、64、32. 参数的优化主要采用自适应矩估

计(Adam)算法,该算法是随机梯度下降算法的一种变形,可以保证每次迭代参数的学习率具有一定的范围.

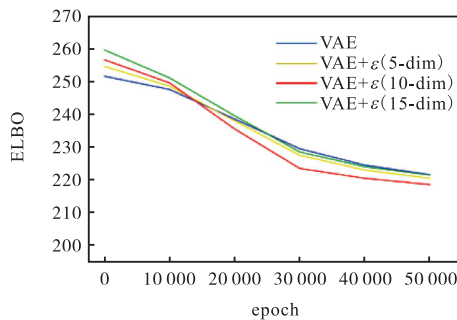
3.2 对比实验

3.2.1 随机噪声维度的选择

通过改变推理模型中随机噪声的维度进行实验,衡量标准为模型的变分似然下界(ELBO).实验中所有模型的隐变量维度均为 10,将注入随机噪声的变分自编码模型的噪声维度分别设置为 5、10、15.模型在两个数据集上的实验结果如图 6 所示,其中 dim 为隐变量维度.根据图 6 的对比结果可知,向推理模型中注入随机噪声可以提高模型的精度,即得到更优的变分似然下界.同时,随机噪声维度不宜设置过大,其维度为 10 时在两个数据集上均得到了更好的实验结果.



(a) MNIST 数据集



(b) Fashion-MNIST 数据集

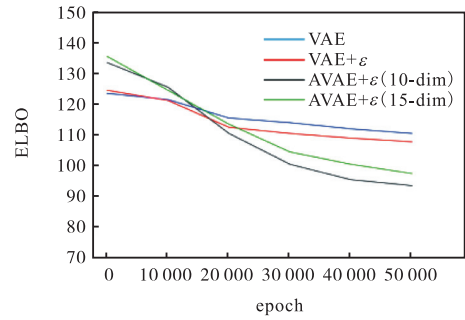
图 6 VAE 模型与注入噪声的 VAE 模型的实验结果

Fig. 6 Experimental results of VAE and injected noise VAE

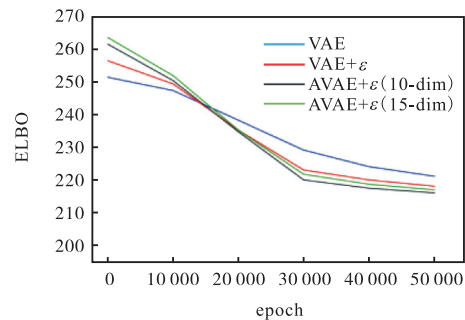
3.2.2 精度对比

通过实验对比 VAE 模型、加入随机噪声的 VAE 模型及 AVAE 模型几种模型的精度,实验的衡量标准为模型的变分似然下界(ELBO).由于随机噪声维度为 10 时实验结果较好,所以设置所有模型的随机噪声维度均为 10,只改变 AVAE 模型的隐变量维度为 10 和 15.模型在两个数据集上的实验结果如图 7 所示,其中 dim 为隐变量维度.根据图 7 的对比结果可

知,AVAE 模型在两个数据集下均得到了更好的实验结果,即得到了更优的变分似然下界.而且相比于改变随机噪声维度,改变隐变量维度对模型结果影响更为显著.与注入随机噪声的变分自编码模型相比,AVAE 模型因为引入对抗策略,得到了更优的变分似然下界;同时,当隐变量维度为 10 时,模型的精度最高.



(a) MNIST 数据集



(b) Fashion-MNIST 数据集

图 7 改变隐变量维度 AVAE 与 VAE 模型的实验结果

Fig. 7 Experimental results of changing hidden variable dimension in AVAE and VAE models

除此之外,将 AVAE 模型与对抗变分贝叶斯模型进行对比,在确保网络结构相同的情况下,随机噪声维度均设置为 10,并改变隐变量维度.实验结果取多次实验精度的平均值,模型在两个数据集上的实验结果见表 1,其中 dim 为隐变量维度.根据表 1 的对比结果可以得到,在推理模型中加入随机噪声的变分自编码模型在两个数据集上均可以提升模型精度,且在隐变量维度为 10 时,得到了更好的实验结果.同时,AVAE 模型因为加入对抗网络,比仅加入随机噪声的变分自编码模型在任何隐变量维度上都得到更高的模型精度.通过对比表 1 的最后两行可以发现,在隐变量为 10 和 15 的情况下,AVAE 模型在两个数据集上均得到了更好的实验结果,即更优的变分似然下界.随着隐变量维度的增加,推理模型需要学习更多的图像特征,使得每一种模型的精度逐渐变差,且当

隐变量维度为 10 时, AVAE 模型在所有模型中得到 的精度最高.

表 1 精度对比实验结果

Tab. 1 Experiment results of accuracy comparison

模型	MNIST			Fashion-MNIST		
	10-dim	15-dim	20-dim	10-dim	15-dim	20-dim
VAE	109.8	111.2	112.8	221.1	223.5	225.4
VAE + ϵ	107.3	109.4	111.1	220.1	222.4	223.9
AVB	95.12	97.01	97.81	216.9	217.3	220.3
AVAE	93.08	96.90	97.90	216.1	217.1	220.6

3.3 生成图像效果

引入随机噪声的变分自编码模型、对抗变分贝叶斯模型和 AVAE 模型的隐变量维度和随机噪声维度均为 10, 生成图像效果如图 8 所示. 相比于引入随机噪声的变分自编码模型, AVAE 模型和对抗变分贝叶

斯模型可以生成更加清晰、真实的图像, 如图 8 中的数字“6”和“手提包”. 实验结果表明对抗网络和聚合正则化约束的引入可以提高模型的精度, 同时提升图像的生成效果.

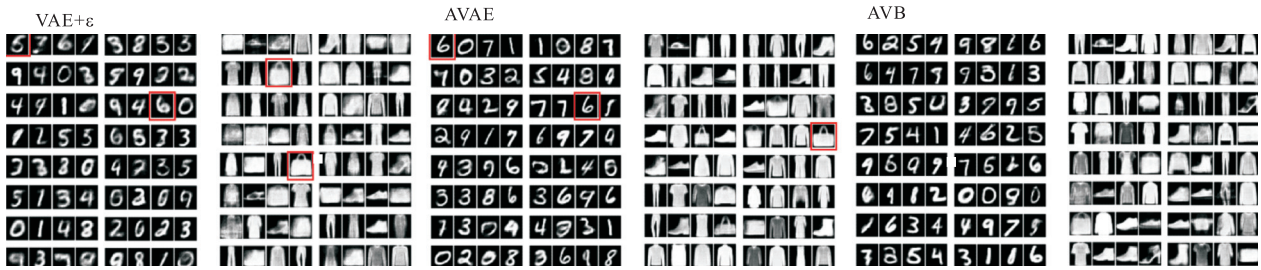


图 8 模型生成图像效果展示
Fig. 8 Model generation effect display

4 结 语

本文提出了一种基于对抗策略的变分自编码模型, 并在实验中展示其模型精度及生成图像效果. 与经典的 3 种模型的对比实验结果显示, 该模型在给定参数的情况下可以得到优于其他模型的变分似然下界(ELBO), 且在 MNIST 和 Fashion-MNIST 数据集上获得了较好的生成图像效果.

经过证实, 在推理模型中加入随机噪声可以提高变分自编码模型的模型精度; 对抗网络和聚合正则化约束的加入均可以提升推理模型的代表能力. 相比于其他模型, 基于对抗策略的变分自编码模型可以提高图像生成的效果, 且在无监督学习中的表现较为突出.

参考文献:

[1] KINGMA D P, WELLING M. Auto-encoding variational Bayes[EB/OL]. [2021-03-30]. <http://arxiv.org/abs/1312.6114>.

[2] DOERSCH C. Tutorial on variational autoencoders

[EB/OL]. [2021-03-30]. <http://arxiv.org/abs/1606.05908>.

[3] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[EB/OL]. [2021-03-30]. <http://arxiv.org/pdf/1406.2661v1.pdf>.

[4] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[EB/OL]. [2021-03-30]. <http://arxiv.org/abs/1511.06434>.

[5] 刘遵雄, 蒋中慧, 任行乐. 多尺度生成对抗网络的图像超分辨率算法[J]. 科学技术与工程, 2020, 20(13): 5217-5223.

[6] 魏媛媛, 倪建成, 高峰, 等. 结合主题信息聚类编码的文本摘要模型[J]. 计算机技术与发展, 2021, 31(1): 30-34.

[7] MAKHZANI A, SHLENS J, JAITLY N, et al. Adversarial autoencoders[EB/OL]. [2021-03-30]. <http://arxiv.org/abs/1511.05644>.

[8] MESCHEDER L, NOWOZIN S, GEIGER A. Adversarial variational Bayes: unifying variational autoencoders and generative adversarial networks[EB/OL]. [2021-03-30]. <http://arxiv.org/abs/1701.04722>.

[9] 殷瑞刚, 魏帅, 李晗, 等. 深度学习中的无监督学习方法综述[J]. 计算机系统应用, 2016, 25(8): 1-7.

- [10] MANUEL L, BELEN C, ANTONIO S, et al. Conditional variational autoencoder for prediction and feature recovery applied to intrusion detection in IoT[J]. *Sensors*, 2017, 17(9): 1967.
- [11] LOUIZOS C, SWERSKY K, LI Y, et al. The variational fair autoencoder[EB/OL]. [2021-03-30]. <http://arxiv.org/abs/1511.00830v1>.
- [12] VINCENT P, LAROCHELLE H, BENGIO Y, et al. Extracting and composing robust features with denoising autoencoders[C]// *Machine Learning, Proceedings of the Twenty-Fifth International Conference (ICML 2008)*. Amsterdam: Elsevier, 2008.
- [13] KINGMA D P, CHEN X, SALIMANS T, et al. Variational lossy autoencoder[EB/OL]. [2021-03-30]. <http://arxiv.org/pdf/1611.02731.PDF>.
- [14] VAN DEN OORD A, NAL K, KORAY K. Pixel recurrent neural networks[EB/OL]. [2021-03-30]. <https://arxiv.org/abs/1601.06759>.
- [15] MIRZA M, OSINDERO S. Conditional generative adversarial nets[EB/OL]. [2021-03-30]. <https://arxiv.org/abs/1411.1784>.
- [16] LARSEN A B L, SØNDERBY S K, WINTHER O. Autoencoding beyond pixels using a learned similarity metric[EB/OL]. [2021-03-30]. <https://arxiv.org/abs/1512.09300>.
- [17] DUMOULIN V, BELGHAZI I, POOLE B. Adversarially learned inference[EB/OL]. [2021-03-30]. <http://arxiv.org/abs/1606.00704>.
- [18] HUANG H, LI Z, HE R, et al. IntroVAE: introspective variational autoencoders for photographic image synthesis[EB/OL]. [2021-03-30]. <http://arxiv.org/pdf/1807.06358.pdf>.
- [19] REZENDE D J, MOHAMED S, WIERSTRA D. Stochastic backpropagation and approximate inference in deep generative models[EB/OL]. [2021-03-30]. <http://www.arxiv.org/pdf/1401.4082.pdf>.

责任编辑: 郎婧

(上接第 55 页)

- 研究[J]. *信息通信*, 2020(12): 152-153.
- [7] LIU B D. *Uncertainty theory*[M]. Berlin: Springer-Verlag, 2007.
- [8] LIU Y, QU Z G, LI X Z, et al. Reliability modelling for repairable systems with stochastic lifetimes and uncertain repair times[J]. *IEEE Transactions on fuzzy systems*, 2019, 27(12): 2396-2405.
- [9] LIU Y, ZHAO J Y, QU Z G, et al. Structural reliability assessment based on subjective uncertainty[EB/OL]. [2021-08-20]. <https://doi.org/10.1142/S0219876221500468>.
- [10] GAO Y, YANG L, LI S, et al. On distribution function of the diameter in uncertain graph[J]. *Information sciences: an international journal*, 2015, 296: 61-74.
- [11] YAO K, LIU B D. Uncertain regression analysis: an approach for imprecise observations[J]. *Soft computing*, 2018, 22: 5579-5582.
- [12] YANG X F, LIU B D. Uncertain time series analysis with imprecise observations[J]. *Fuzzy optimization & decision making*, 2018, 18: 263-278.
- [13] 刘博宁, 张鹏, 张建业, 等. 软件可靠度的模糊故障树评定方法[J]. *计算机应用研究*, 2012, 29(10): 3783-3786.

责任编辑: 郎婧