

弱水三千，只取你标！AL（主动学习）结合GAN如何？

原创 bryant8 机器学习与生成对抗网络 2019-12-24

欢迎点击上方蓝字，关注啦~

相关阅读：

GAN整整6年了！是时候要来捋捋了！

异常检测，GAN如何gan？

虚拟换衣！速览这几篇最新论文咋做的！

脸部妆容迁移！速览几篇用GAN来做的论文

【1】GAN在医学图像上的生成，今如何？

01-GAN公式简明原理之铁甲小宝篇

这次简单记录下、GAN和主动学习结合的一些论文，不当处、望指正~

Active Learning 主动学习：

背景

众所周知，深度学习的崛起和广泛应用是依靠着大量的标注数据的，但在很多场合下，大规模数据的标注成本太高，同时也可能导致训练时间过长。主动学习可挑出所谓高信息的数据去标注，从而降低标注成本、减少训练时间，还可以迭代提升模型表现。

定义

目的是设计一个选择/查询函数（query function），用它来从大量的、未标注的数据池中选出具有高价值的待标注数据，递送给人工标注（oracle）后，加入训练集，反复迭代训练模型。

常见手段

主动方法常见的有基于池、基于合成的方法。

- **基于池(pool)**：根据预设的选择策略选出的数据交给基准分类器预测，错误时再送人工标注。对于查询策略，如何挑选最有信息量的样本，常见地：
 - 1) Random Sampling：随机选择；
 - 2) Uncertainty Sampling：选择当前模型最不确定的样本，如分类概率为0.5等。但显然，这种策略受异常点、outlier 样本、冗余的样本影响。
- **基于合成**：使用生成模型生成更具有信息的样本。



1. 2017-Generative Adversarial Active Learning

<https://arxiv.xilesou.top/pdf/1702.07956.pdf>

简介：

第一个将GAN结合主动学习的工作，提出GAAL。采用的手段是**基于合成**的思路。如下图所示，还是一目了然的。

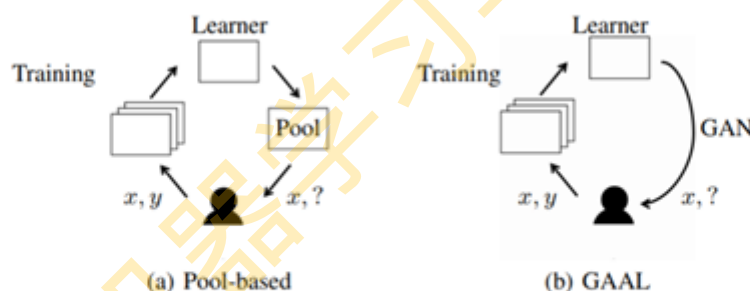


Figure 1: (a) Pool-based active learning scenario. The learner selects samples for querying from a given unlabeled pool. (b) GAAL algorithm. The learner synthesizes samples for querying using GAN.

The main contributions of this work are as follows:

- To the best of our knowledge, this is the first active learning framework using deep generative models¹.
- While we do not claim our method is always superior to the previous active learners in terms of accuracy, in some cases, it yields classification performance not achievable even by a fully supervised learning scheme. With enough capacity from the trained generator, our method allows us to have control over the generated instances which may not be available to the previous active learners.

有趣的是，作者在贡献的声明强调，该工作贡献主要在于：首次将GAN和主动学习相结合，而不是要和各种SOTA方法在什么精确率上一较高下，所以论文的实验部分，作者采用的是SVM做为分类器。

The main contributions of this work are as follows:

- To the best of our knowledge, this is the first active learning framework using deep generative models¹.
- While we do not claim our method is always superior to the previous active learners in terms of accuracy, in some cases, it yields classification performance not achievable even by a fully supervised learning scheme. With enough capacity from the trained generator, our method allows us to have control over the generated instances which may not be available to the previous active learners.

当然，作者也说，我们方法是极具前景的~，并且和基于池的相比，也是competitive的呢，我们的方法也许可以启发后来者、用GAN展开相关的工作（坑已挖好，来跳，哈哈）。

- We conduct experiments to compare our active learning approach with self-taught learning². The results are promising.
- This is the first work to report numerical results in active learning synthesis for image classification. See [43, 30]. The proposed framework may inspire future GAN applications in active learning.
- The proposed approach should not be understood as a pool-based active learning method. Instead, it is active learning by query synthesis. We show that our approach can perform competitively when compared against pool-based methods.

实验：

对于训练数据的初始化是，随机筛选50个样本；每次1个batch对应10次queries。进行了以下几种方法的对比实验：

The following schemes are implemented and compared in our experiments.

- The proposed generative adversarial active learning (GAAL) algorithm as in Algorithm 1.
- Using regular GAN to generate training data. We refer to this as simple GAN.
- SVM_{active} algorithm from [45].
- Passive random sampling, which randomly samples instances from the unlabeled pool.
- Passive supervised learning, i.e., using all the samples in the pool to train the classifier.
- Self-taught learning from [39].

作者进行的是二分类的实验（多分类类似），MNIST上对数字5和7分类，CIFAR10对 automobile 和 horse（也许因为当时GAN生成能力有限，或者也许作者懒得去搞最SOTA的GAN了，反正能够说明问题就行。但是采用的DCGAN实在乏力，在其实验数据上连生成猫和狗都吃力，因此选汽车和马生成效果更好区分些，2333）。作者其实也有提到一些可以改进GAN的方法，但作为未来工作，现在的实验结果先发文发出来，哈哈。

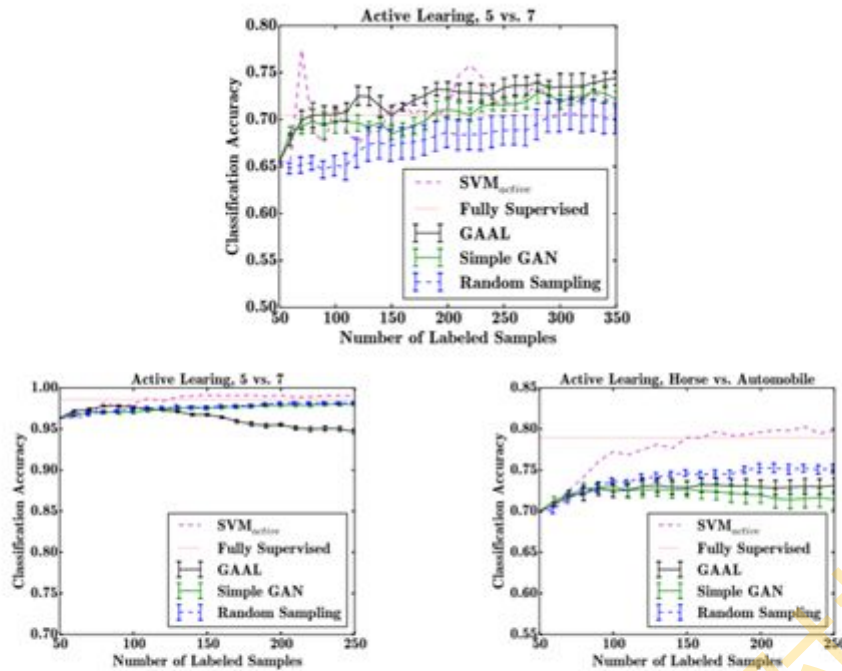


Figure 5: Active learning results. (Top) Train on MNIST, test on USPS, Classifying 5 and 7. The results are averaged over 10 runs. (Bottom Left) Train on MNIST, test on MNIST, Classifying 5 and 7. (Bottom Right) CIFAR-10 dataset, classifying automobile and horse. The results are averaged over 10 runs. The error bars represent the empirical standard deviation of the average values. The figures are best viewed in color.

在上图的top子图可以看到，在350个训练样本的时候，GAAL开始超越SVM-active和全监督训练的方法。其他图的具体细节感兴趣可以阅读原文。更多实验结果还有：

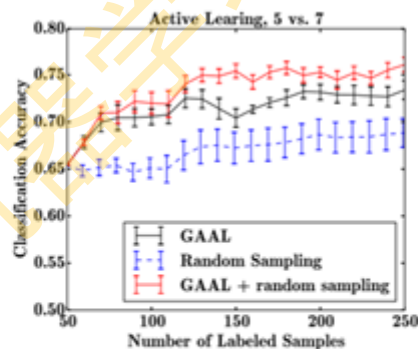


Figure 6: Active learning results using a mixed scheme. The mixed scheme executes one iteration of random sampling after every five iterations of GAAL algorithm. Train on MNIST, test on USPS, Classifying 5 and 7. The results are averaged over 10 runs. The error bars represent the empirical standard deviation of the average values. The figure is best viewed in color.

2. 2019-10-28 Variational Adversarial Active Learning

<https://arxiv.xilesou.top/pdf/1904.00370.pdf>

简介：

提出一种**基于池**的半监督主动学习算法，通过对抗的方式学习采样/选择机制。

使用变分自编码器学习潜码空间，训练对抗网络的判别器区分数据是否被标记。进一步地，VAE和判别器之间进行对抗学习：

VAE尽可能让判别器预测所有的数据都是来自于标记池，判别器尽可能在隐空间层面区分是否为标记数据。作者认为所提出的方法可以学习有效的、低维的隐空间**表征**，并提供了一种高效的**采样/选择方法**。

如下图所示， (X_L, Y_L) 表示打好标签的**标记池**中的标记数据， (X_U) 表示在大量未标记数据池中的数据。目标是训练最label-efficient的模型：通过迭代地查询一个固定的采样预算，从未标记池中挑选出最有“信息价值”的b个样本，提供给oracle（人工）进行标注。

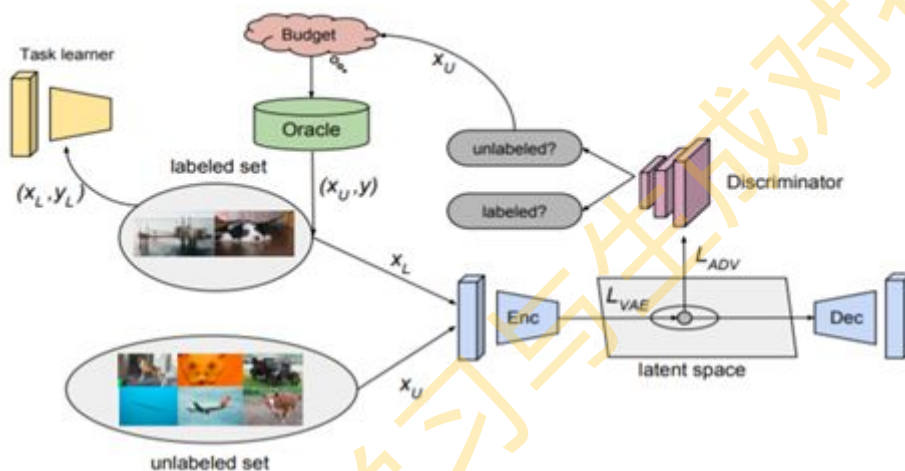


Figure 1. Our model (VAAL) learns the distribution of labeled data in a latent space using a VAE optimized using both reconstruction and adversarial losses. A binary adversarial classifier (discriminator) predicts unlabeled examples and sends them to an oracle for annotations. The VAE is trained to fool the adversarial network to believe that all the examples are from the labeled data while the discriminator is trained to differentiate labeled from unlabeled samples. Sample selection is entirely separate from the main-stream task for which we are labeling data inputs, making our method to be *task-agnostic*

表征学习：

Transductive representation learning

使用 β 变分自编码器进行表征学习。编码器在高斯先验基础上，学习一个隐含低维表征空间，解码器可以重建输入的数据。为补全有标记的数据在表征学习过程中丢失的特征，采用的是**transductive learning**（在训练过程中，已知testing data (unlabelled data)）：

$$\mathcal{L}_{\text{VAE}}^{\text{trd}} = \mathbb{E}[\log p_{\theta}(x_L|z_L)] - \beta \text{D}_{\text{KL}}(q_{\phi}(z_L|x_L)||p(z)) \\ + \mathbb{E}[\log p_{\theta}(x_U|z_U)] - \beta \text{D}_{\text{KL}}(q_{\phi}(z_U|x_U)||p(z)) \quad (1)$$

where q_{ϕ} and p_{θ} are the encoder and decoder parameterized by ϕ and θ , respectively. $p(z)$ is the prior chosen as a unit Gaussian, and β is the Lagrangian parameter for the optimization problem. The reparameterization trick is used for proper calculation of the gradients [28].

Adversarial representation learning

前面说过，大多数的采样策略的根据是模型的不确定性，例如认为：模型对预测越不确定，未标记样本包含的信息越多。但这种方法受限于异常点。相反，此方法对于采样策略的处理是，通过训练对抗网络去学习如何区分在潜在空间的表征。对抗网络中将输入映射到潜码空间，并且给一个标签，若样本是标记数据，则为1，如果是未标记数据，则为0。关键是、使用对抗的方式，VAE将标记和未标记的数据都映射到相似概率分布的空间，去欺骗判别器说所有的输入均是标记的。当然，判别器则尝试避免欺骗：

$$\mathcal{L}_{\text{VAE}}^{\text{adv}} = -\mathbb{E}[\log(D(q_{\phi}(z_L|x_L)))] - \mathbb{E}[\log(D(q_{\phi}(z_U|x_U)))] \quad (2)$$

$$\mathcal{L}_D = -\mathbb{E}[\log(D(q_{\phi}(z_L|x_L)))] - \mathbb{E}[\log(1 - D(q_{\phi}(z_U|x_U)))] \quad (3)$$

$$\mathcal{L}_{\text{VAE}} = \lambda_1 \mathcal{L}_{\text{VAE}}^{\text{trd}} + \lambda_2 \mathcal{L}_{\text{VAE}}^{\text{adv}} \quad (4)$$

Algorithm 1 Variational Adversarial Active Learning

Input: Labeled pool (X_L, Y_L) , Unlabeled pool (X_U) , Initialized models for θ_T , θ_{VAE} , and θ_D

Input: Hyperparameters: epochs, λ_1 , λ_2 , α_1 , α_2 , α_3

```

1: for  $e = 1$  to epochs do
2:   sample  $(x_L, y_L) \sim (X_L, Y_L)$ 
3:   sample  $x_U \sim X_U$ 
4:   Compute  $\mathcal{L}_{VAE}^{trd}$  by using Eq. 1
5:   Compute  $\mathcal{L}_{VAE}^{adv}$  by using Eq. 2
6:    $\mathcal{L}_{VAE} \leftarrow \lambda_1 \mathcal{L}_{VAE}^{trd} + \lambda_2 \mathcal{L}_{VAE}^{adv}$ 
7:   Update VAE by descending stochastic gradients:
8:    $\theta'_{VAE} \leftarrow \theta_{VAE} - \alpha_1 \nabla \mathcal{L}_{VAE}$ 
9:   Compute  $\mathcal{L}_D$  by using Eq. 3
10:  Update  $D$  by descending its stochastic gradient:
11:   $\theta'_D \leftarrow \theta_D - \alpha_2 \nabla \mathcal{L}_D$ 
12:  Train and update  $T$ :
13:   $\theta'_T \leftarrow \theta_T - \alpha_3 \nabla \mathcal{L}_T$ 
14: end for
15: return Trained  $\theta_T, \theta_{VAE}, \theta_D$ 

```

采样/选择策略:

假如说要挑选 b 个高质量样本给人工标注，所用依据是鉴别器的预测分数（挑选 b 个最低的自信度， D 判断出来越小的，越可能是未标记池中的数据）。

Algorithm 2 Sampling Strategy in VAAL

Input: b, X_L, X_U

Output: X_L, X_U

```

1: Select samples  $(X_s)$  with  $\min_b \{\theta_D(z_U)\}$ 
2:  $Y_o \leftarrow \mathcal{ORACLE}(X_s)$ 
3:  $(X_L, Y_L) \leftarrow (X_L, Y_L) \cup (X_s, Y_o)$ 
4:  $X_U \leftarrow X_U - X_s$ 
5: return  $X_L, X_U$ 

```

实验:

作者在分类分割等任务都做了许多实验、去验证所提出的方法的有效性，这里就不贴了，详见原论文。

3. 2019-12-20 Adversarial Representation Active Learning

简介：

GAAL严重依赖于生成图像的质量，并且生成器和鉴别器并没有得到迭代式的提升。而VAAL仅使用了标注数据训练分类器。不同之前的方法，该文作者在VAAL基础上，提出不仅使用已标注的数据训练分类器，同时还使用未标注的和生成的数据联合训练整个模型。

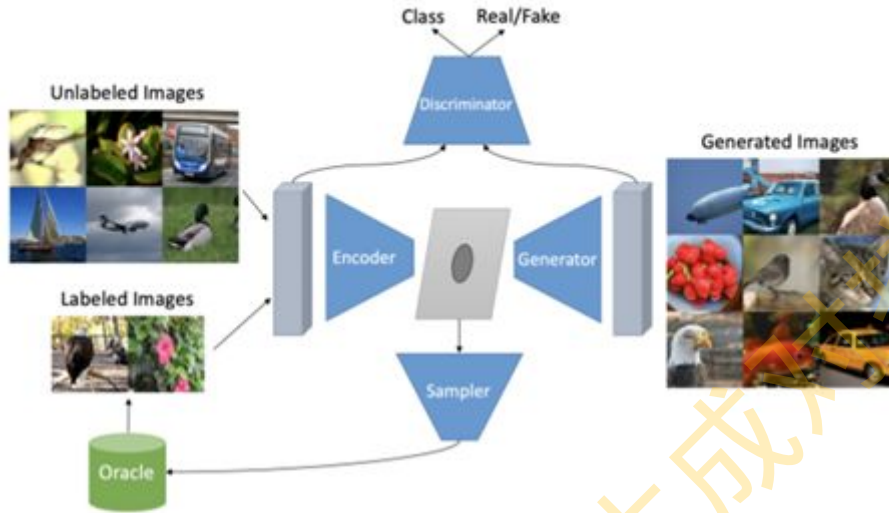


Figure 1: At each iteration, our model learns a latent representation of both labeled and unlabeled images and simultaneously trains an efficient classifier using labeled and class-conditional generated images. Then, it selects the most informative unlabeled images (based on their latent representation) to be labeled by the oracle for the next iteration.

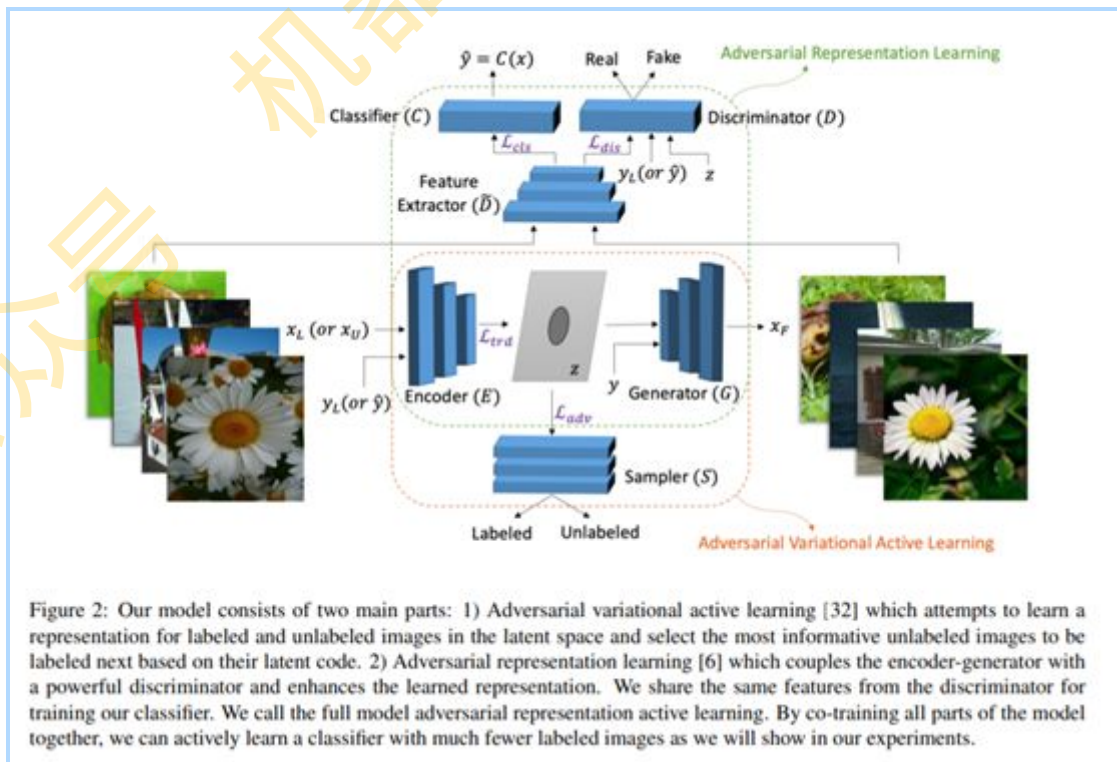



Figure 2: Our model consists of two main parts: 1) Adversarial variational active learning [32] which attempts to learn a representation for labeled and unlabeled images in the latent space and select the most informative unlabeled images to be labeled next based on their latent code. 2) Adversarial representation learning [6] which couples the encoder-generator with a powerful discriminator and enhances the learned representation. We share the same features from the discriminator for training our classifier. We call the full model adversarial representation active learning. By co-training all parts of the model together, we can actively learn a classifier with much fewer labeled images as we will show in our experiments.

先暂时写到这吧==

更多分享欢迎关注本公众号：



学点诗歌和AI知识

SCUT - 野  场常年划水者、书法业余爱好者，尝试分享DL、CV、GAN、唐诗宋词三百，好玩、有趣、经典！



让人怪不好意思

机器学习与生成对抗网络

公众号