Semi-supervised learning

**Gapeng-北京大学-CV 2017-07-27 20:00:40**

好了，到点了，我们开始吧，今天的主题是semi-supervised learning

**陈通宝-中科院-机器学习 2017-07-27 20:01:04**

Ok

**harric-西浦-数据生成扩充 2017-07-27 20:01:25**

hao

**harric-西浦-数据生成扩充 2017-07-27 20:01:25**

好

**Gapeng-北京大学-CV 2017-07-27 20:01:25**

用GAN可以怎么做semi-supervised learning？

**郑华滨-中山大学-深度学习 2017-07-27 20:01:35**

总共四篇论文，是一篇篇讨论过去咩

**Gapeng-北京大学-CV 2017-07-27 20:01:50**

都可以

**Gapeng-北京大学-CV 2017-07-27 20:02:02**

也可以方法交叉对比

**harric-西浦-数据生成扩充 2017-07-27 20:02:03**

第一篇是不是dcgan？

**张治坤 浙大 信息安全 2017-07-27 20:02:22**

我之前只看过最后一篇

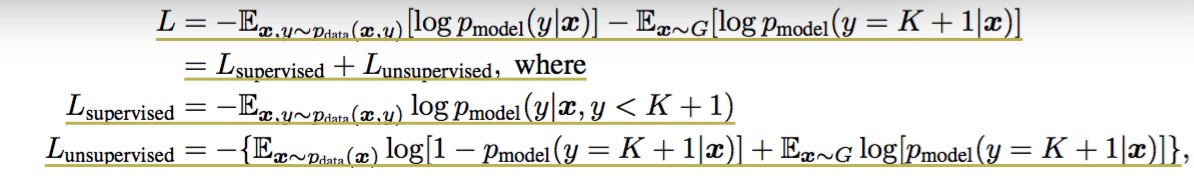
**张治坤 浙大 信息安全 2017-07-27 20:02:40**

发在NIPS2016上的

**ccm-GZ-ML 2017-07-27 20:02:47**

improved GAN里是把loss 改成

**ccm-GZ-ML 2017-07-27 20:02:47**



**张治坤 浙大 信息安全 2017-07-27 20:02:53**

坐等大佬们讲解

**Gapeng-北京大学-CV 2017-07-27 20:02:54**

那你可以讲讲improved GAN怎么做ssl的

**郑华滨-中山大学-深度学习 2017-07-27 20:03:04**

我也只看过最后一篇，要不要让读过这四篇的人先分别概述一下每篇的思想

**林同学-西工大-CV 2017-07-27 20:03:17**

可以！

**harric-西浦-数据生成扩充 2017-07-27 20:03:17**

题外话 今天楼主会不会整理笔记呀？

**Gapeng-北京大学-CV 2017-07-27 20:03:26**

会

**harric-西浦-数据生成扩充 2017-07-27 20:03:37**

之前几次讨论我都在国外 你们讨论的时候我在睡觉 都只能后来看笔记 今天回国啦哈哈哈

**harric-西浦-数据生成扩充 2017-07-27 20:03:39**

好滴！

**Gapeng-北京大学-CV 2017-07-27 20:04:14**

所以群里有没有看过四篇的？

**harric-西浦-数据生成扩充 2017-07-27 20:04:23**

其实我都看过

**兔子-不存在-ML 2017-07-27 20:04:33**

那你说说呗

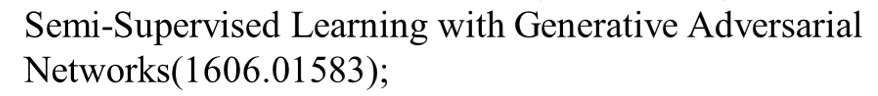
**Gapeng-北京大学-CV 2017-07-27 20:04:39**

你来讲讲吧

**harric-西浦-数据生成扩充 2017-07-27 20:04:47**

这一个的方法是不是后来发展成了infogan还是improvedgan？

**harric-西浦-数据生成扩充 2017-07-27 20:04:51**



**harric-西浦-数据生成扩充 2017-07-27 20:04:54**

我都忘记了。。。

**郑华滨-中山大学-深度学习 2017-07-27 20:04:57**

对呀几段话概述一下[奸笑]

**harric-西浦-数据生成扩充 2017-07-27 20:05:00**

[委屈][委屈][委屈]

**harric-西浦-数据生成扩充 2017-07-27 20:05:22**

我记得这个是一个很短的 作者最后写了一句 有人用了同样的思想blablabla

**hinse-scnu-nlp 2017-07-27 20:05:23**

"第一篇感觉是VAE 的方法  
"

**郑华滨-中山大学-深度学习 2017-07-27 20:05:37**

感觉improved gan这篇比较经典，读过的人比较多

**Gapeng-北京大学-CV 2017-07-27 20:05:42**

对的，第一篇是VAE的

**harric-西浦-数据生成扩充 2017-07-27 20:05:52**

噢对 第一篇是vae 我想起来了

**hinse-scnu-nlp 2017-07-27 20:06:00**

第一篇对照vae就很好理解了

**hinse-scnu-nlp 2017-07-27 20:07:00**

半监督就是把y 也作为一个latent , 和z一起

**Clockworkai\_西电\_NLP 2017-07-27 20:07:34**

唉，错过了seqGAN的讨论

**lm\_医渡云\_识别 2017-07-27 20:08:05**

很想知道用GAN类型的算法做ssl有什么优点

**郑华滨-中山大学-深度学习 2017-07-27 20:08:05**

"思想总共有K类真实图片，然后让判别器预测K+1个类，其中多出来的一个类是表示假样本。  
  
对于标注了类别k的真实样本，判别器要给出第k个概率高  
对于未标注真实样本，判别器要给出第k+1个概率低，即表示不为假，但是不知道具体是哪一类  
对于生成样本，判别器要给出第k+1个概率高，即表示为假"

**林同学-西工大-CV 2017-07-27 20:08:10**

有记录可以看

**Gapeng-北京大学-CV 2017-07-27 20:08:50**

Improved GAN的D是个分类器，分成c+1类，G生成的数据作为多出来的那一类，而那c类是正样本。第一篇SSL其实是VAE的做法，M1模型学特征到图像的分布，M2学latent space到特征的分布，M1+M2实现完整的分类。

**黄瑞阳\_郑州大学\_NLP 2017-07-27 20:09:01**

今天不是讲半监督的GAN吗？一下子就讲Improved GAN了？

**hinse-scnu-nlp 2017-07-27 20:09:13**

"好像也可以知道是哪一类，D如果觉得不u假， 也要让C归类正确  
"

**郑华滨-中山大学-深度学习 2017-07-27 20:09:14**

improved gan有做半监督

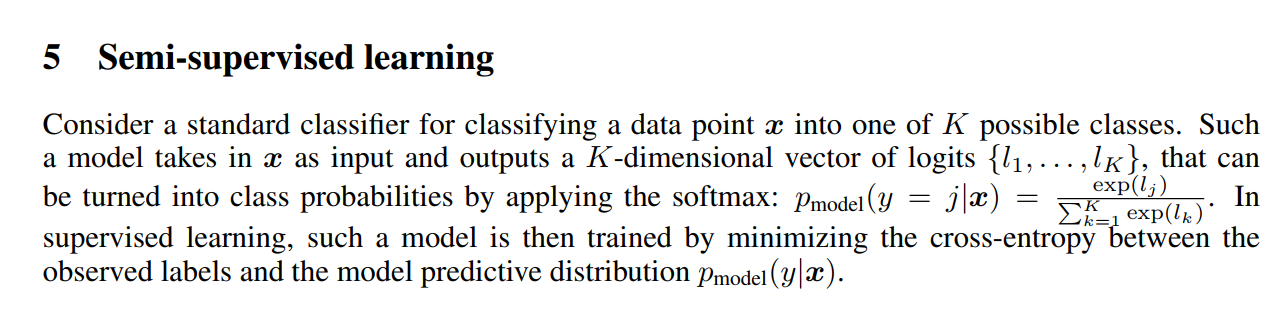
**郑华滨-中山大学-深度学习 2017-07-27 20:09:21**

没毛病

**ccm-GZ-ML 2017-07-27 20:09:21**

improved GAN里面也有ssl部分...

**张治坤 浙大 信息安全 2017-07-27 20:10:07**



**张治坤 浙大 信息安全 2017-07-27 20:10:10**

对，第五部分

**郑华滨-中山大学-深度学习 2017-07-27 20:10:30**

感觉可以先讨论第四篇，读的人比较多

**郑华滨-中山大学-深度学习 2017-07-27 20:10:37**

就improved

**Gapeng-北京大学-CV 2017-07-27 20:10:41**

嗯

**Gapeng-北京大学-CV 2017-07-27 20:10:49**

就从improved gan开始吧

**郑华滨-中山大学-深度学习 2017-07-27 20:10:55**

我想先提一个问题，为什么这样训练就能分对未标注样本

**hinse-scnu-nlp 2017-07-27 20:10:58**

恩 这边是比较出名

**郑华滨-中山大学-深度学习 2017-07-27 20:11:43**

我在网上搜过一些讨论，没有讲得比较清楚的，为什么它这样设计就能ssl

**ccm-GZ-ML 2017-07-27 20:13:06**

有人用Improved gan ssl的框架做过吗？

**陆鹏起-hust-gan 2017-07-27 20:13:25**

SSL涉及三种误差。

**陆鹏起-hust-gan 2017-07-27 20:13:37**

1.对于训练集中的有标签样本，考察估计的标签是否正确。即，计算分类为相应的概率

**陆鹏起-hust-gan 2017-07-27 20:13:55**

2.对于训练集中的无标签样本，考察是否估计为“真”。即，计算不估计为K+1类的概率

**陆鹏起-hust-gan 2017-07-27 20:14:11**

3.对于生成器产生的伪样本，考察是否估计为“伪”。即，计算估计为K+1类的概率

**Gapeng-北京大学-CV 2017-07-27 20:14:18**

它这里的半监督的意思只是因为多用了G生成的数据？而G生成的数据是不知道label的

**hinse-scnu-nlp 2017-07-27 20:14:21**

"Discriminator perspective. The requirements to the discriminator are that it should (i) be certain  
of class assignment for samples from D, (ii) be uncertain of assignment for generated samples, and  
(iii) use all classes equally 3 .  
Generator perspective. The requirements to the generator are that it should (i) generate samples  
with highly certain class assignments, and (ii) equally distribute samples across all K classes.  
"

**anshiquanshu66-师大-医学图像处理 2017-07-27 20:14:58**

SSL是啥？表示不懂

**周昕宇Tim-旷视科技-深度学习 2017-07-27 20:15:09**

semi supervised learning

**深度挖掘机-NUS-NLP 2017-07-27 20:15:13**

semi supervised learning

**张治坤 浙大 信息安全 2017-07-27 20:15:13**

Semi-supervised learning

**侯瑞兵-计算所-计算机视觉 2017-07-27 20:15:28**

本身就是半监督的

**郑华滨-中山大学-深度学习 2017-07-27 20:15:40**

generate samples with highly certain class assignment这点哪里有体现？

**Gapeng-北京大学-CV 2017-07-27 20:15:56**

训练集如果有无标签的数据，这里怎么做？

**郑华滨-中山大学-深度学习 2017-07-27 20:16:05**

生成器目标仅仅是要生成样本不会被分到假而已啊

**黄瑞阳\_郑州大学\_NLP 2017-07-27 20:16:26**

有人比较过吗？现在这种GAN的分类效果怎样？

**Gapeng-北京大学-CV 2017-07-27 20:16:39**

这就需要依靠D的分类性能了

**侯瑞兵-计算所-计算机视觉 2017-07-27 20:17:00**

是这样的。。。所以其实判别器始终不会对无标签的数据进行分类。。

**陆鹏起-hust-gan 2017-07-27 20:17:21**

生成器的损失函数跟原始gan是一样的。。只是希望生成的样本骗过分类器而已

**Gapeng-北京大学-CV 2017-07-27 20:17:23**

也就是说D训练的好，分类越准确，G也就越能生成highly certain class

**彭彭-浙工大-主动学习 2017-07-27 20:17:27**

无标签的数据被D分出来了吧

**郑华滨-中山大学-深度学习 2017-07-27 20:17:31**

训练集无标签样本要被判别器分到前K类的任何一个，所以反过来只要它别被判别器分到第K+1类即可

**lm\_医渡云\_识别 2017-07-27 20:18:00**

我以前做ssl是根据有标签的数据，把标签扩散到无标签数据。这里说的好像目标都不一样吧？

**郑华滨-中山大学-深度学习 2017-07-27 20:18:00**

没有highly certain class

**郑华滨-中山大学-深度学习 2017-07-27 20:18:00**

没有体现这一点

**hinse-scnu-nlp 2017-07-27 20:18:06**

"没标签只能辨别真假， 不能给你辨别到哪一类  
"

**张治坤 浙大 信息安全 2017-07-27 20:18:11**

那这样的话如何提高原始分类器的精度呢

**Gapeng-北京大学-CV 2017-07-27 20:18:20**

我觉得在这种训练模式下，训练集不能有无标签样本

**hinse-scnu-nlp 2017-07-27 20:18:30**

@ 华滨 我贴的那个不是对imporoved gan

**郑华滨-中山大学-深度学习 2017-07-27 20:18:33**

标签扩散只是ssl的一种范式而已，我们这里讨论的是半监督，不仅限于标签扩散的范式

**林同学-西工大-CV 2017-07-27 20:18:35**

分类的话 还是不好做吧Z 。一般resnet这些的 还是好调试感觉

**张治坤 浙大 信息安全 2017-07-27 20:18:42**

和监督学习比起来他的优点在哪，这个我还是很懂

**郑华滨-中山大学-深度学习 2017-07-27 20:18:42**

哦哦好

**hinse-scnu-nlp 2017-07-27 20:18:44**

improved gan 无监督只能分真假

**侯瑞兵-计算所-计算机视觉 2017-07-27 20:19:40**

这样分类精度会提高吗

**lm\_医渡云\_识别 2017-07-27 20:19:40**

所以只是说训练数据中一部分有标签，一部分无标签，无标签的部分可以帮助提高分类器效果？

**子棐-安防-学习 2017-07-27 20:19:41**

那无标签这部分是如何实现的呢？

**郑华滨-中山大学-深度学习 2017-07-27 20:19:52**

对的，提高分类器效果

**Gapeng-北京大学-CV 2017-07-27 20:20:05**

所以是不是可以认为，它不是真正的ssl？

**hinse-scnu-nlp 2017-07-27 20:20:19**

"因为训练集中标注工作量大  
"

**郑华滨-中山大学-深度学习 2017-07-27 20:20:19**

"总共有K类真实图片，然后让判别器预测K+1个类，其中多出来的一个类是表示假样本。  
  
对于标注了类别k的真实样本，判别器要给出第k个概率高  
对于未标注真实样本，判别器要给出第k+1个概率低，即表示不为假，但是不知道具体是哪一类  
对于生成样本，判别器要给出第k+1个概率高，即表示为假"

**郑华滨-中山大学-深度学习 2017-07-27 20:20:22**

不，它就是

**Gapeng-北京大学-CV 2017-07-27 20:20:37**

不过GAN生成的数据跟原始数据不一样，算是一个数据扩充

**郑华滨-中山大学-深度学习 2017-07-27 20:20:43**

很纯正的ssl，用无标签数据提高分类效果

**郑华滨-中山大学-深度学习 2017-07-27 20:21:01**

不是，生成的数据没有标签，不是数据扩充

**侯瑞兵-计算所-计算机视觉 2017-07-27 20:21:19**

不是很明白为什么可以提高分类精度？

**Gapeng-北京大学-CV 2017-07-27 20:21:25**

训练集上不能有无标签数据吧？

**彭彭-浙工大-主动学习 2017-07-27 20:21:34**

可以有吧

**侯瑞兵-计算所-计算机视觉 2017-07-27 20:21:36**

可以有啊

**郑华滨-中山大学-深度学习 2017-07-27 20:21:43**

所以说improved gan这篇其实没那么简单，好像被误解了挺多[捂脸]

**郑华滨-中山大学-深度学习 2017-07-27 20:21:46**

可以有

**张治坤 浙大 信息安全 2017-07-27 20:21:49**

没有的话就是监督学习了

**张治坤 浙大 信息安全 2017-07-27 20:21:52**

[捂脸]

**林同学-西工大-CV 2017-07-27 20:22:13**

数据集上没有无标签的 那就不能算是ssl了。

**郑华滨-中山大学-深度学习 2017-07-27 20:22:26**

不行我要截一句关键的话

**林同学-西工大-CV 2017-07-27 20:22:26**

有谁做过实验？

**子棐-安防-学习 2017-07-27 20:22:28**

那无标签数据训练过程是怎么样的呢 不是很明白

**郑华滨-中山大学-深度学习 2017-07-27 20:22:29**

This approach introduces an interaction between G and our classiﬁer that we do not fully understand yet, but empirically we ﬁnd that optimizing G using feature matching GAN works very well for semi-supervised learning, while training G using GAN with minibatch discrimination does not work at all.

**郑华滨-中山大学-深度学习 2017-07-27 20:22:47**

improved gan要配合feature matching才能做ssl

**郑华滨-中山大学-深度学习 2017-07-27 20:22:50**

这很关键

**hinse-scnu-nlp 2017-07-27 20:23:02**

"其实这篇文章第一次我只看了那几个技巧  
"

**郑华滨-中山大学-深度学习 2017-07-27 20:23:08**

不配合feature matching的话好像是做不了的

**Gapeng-北京大学-CV 2017-07-27 20:23:22**

嗯，是的，试过不用feature matching，根本不work

**郑华滨-中山大学-深度学习 2017-07-27 20:23:37**

feature matching的意思是，要求真假样本在判别器的中间层feature map的均值要接近

**郑华滨-中山大学-深度学习 2017-07-27 20:23:49**

这就很神奇了，说明关键点可能都不在那三个loss上面，反而在feature matching上面

**Gapeng-北京大学-CV 2017-07-27 20:23:52**

@Xiao-pku-math 做过这方面的实验

**Gapeng-北京大学-CV 2017-07-27 20:24:41**

好，那么问题来了，feature matching为什么能有效果？

**林同学-西工大-CV 2017-07-27 20:24:50**

刚想问Z

**郑华滨-中山大学-深度学习 2017-07-27 20:25:30**

我看到reddit上有一个猜想，说feature matching是把有标注样本和无标注样本的特征空间拉近，对齐

**张治坤 浙大 信息安全 2017-07-27 20:25:36**

这个文章中好像木有讨论

**侯瑞兵-计算所-计算机视觉 2017-07-27 20:25:39**

是不是使真假样本在判别器的feature上尽可能接近。。

**郑华滨-中山大学-深度学习 2017-07-27 20:26:07**

这样如果一个无标注样本a跟一个有标注样本b在特征空间上很接近，那它就会被分为b同样的类

**郑华滨-中山大学-深度学习 2017-07-27 20:26:34**

但是好像又有点说不太通

**林同学-西工大-CV 2017-07-27 20:26:50**

那这样 如果不是同一类的

**Gapeng-北京大学-CV 2017-07-27 20:26:52**

哪里说不通？

**林同学-西工大-CV 2017-07-27 20:27:02**

岂不是错误了？

**侯瑞兵-计算所-计算机视觉 2017-07-27 20:27:08**

但是怎么保证label为a的无标注样本和label为a的有标注样本特征拉近呢

**王强+hdu+ml 2017-07-27 20:27:14**

他怎么知道不是统一类

**林同学-西工大-CV 2017-07-27 20:27:29**

肯定存在这个情况

**郑华滨-中山大学-深度学习 2017-07-27 20:27:38**

因为它只是要拉近真假样本的feature分布，没有说要拉近labelled真与unlabelled真的分布

**hinse-scnu-nlp 2017-07-27 20:28:30**

"感觉G就是要生成中间特征和真比较像的。  
"

**郑华滨-中山大学-深度学习 2017-07-27 20:28:30**

但是似乎在达成前者的过程中有可能会顺便达成后者，不过我也没想明白

**林同学-西工大-CV 2017-07-27 20:28:42**

那为什么选在中间？

**林同学-西工大-CV 2017-07-27 20:28:48**

有点晕

**Gapeng-北京大学-CV 2017-07-27 20:29:06**

那也就是说，只能寄希望于在labeled data上D能够提取到足够好的特征，也就是说，labeled data还是不能太少？

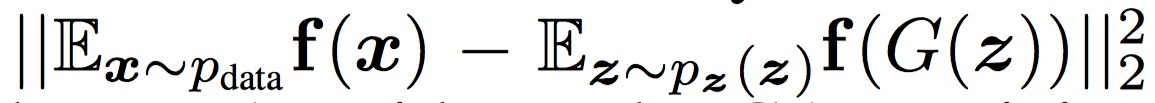
**hinse-scnu-nlp 2017-07-27 20:29:53**

为什么D不能也 feature matching

**郑华滨-中山大学-深度学习 2017-07-27 20:29:58**

又不是，它做mnist只用了100个标注样本，每个数字10个标注样本

**郑华滨-中山大学-深度学习 2017-07-27 20:30:48**



**林同学-西工大-CV 2017-07-27 20:30:51**

这个标注和非标注的数量比例大概多少？

**郑华滨-中山大学-深度学习 2017-07-27 20:30:57**

截个feature matching的目标函数

**郑华滨-中山大学-深度学习 2017-07-27 20:31:21**

除了100个标注，其他mnist都是非标注

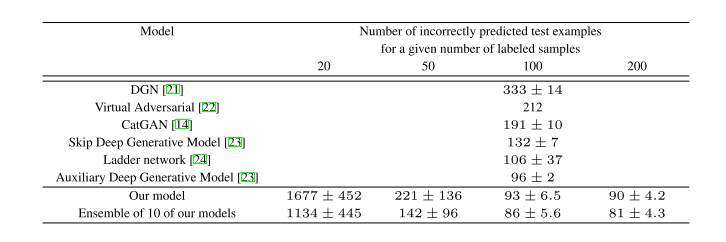
**侯瑞兵-计算所-计算机视觉 2017-07-27 20:31:49**



**林同学-西工大-CV 2017-07-27 20:32:45**

mnist上的分类效果做到了多少？

**Gapeng-北京大学-CV 2017-07-27 20:33:17**



**Ailsa-多伦多大学-生成模型 2017-07-27 20:33:35**

他只给了一些主观误差我记得

**郑华滨-中山大学-深度学习 2017-07-27 20:35:05**

还有一种解释，是原论文给出的，但是我觉得比较玄学：为了区分真假而被迫学到的feature，对于分类也会有用

**Ailsa-多伦多大学-生成模型 2017-07-27 20:35:42**

这种feature matching loss 是不是跟perceptual loss有相似之处呢？只不过一个是用判别网 一个另用了vgg?

**郑华滨-中山大学-深度学习 2017-07-27 20:35:54**

诶，你说对了

**郑华滨-中山大学-深度学习 2017-07-27 20:35:58**

确实很像

**Gapeng-北京大学-CV 2017-07-27 20:35:58**

是的

**Ailsa-多伦多大学-生成模型 2017-07-27 20:36:06**

但都是feature层面的

**郑华滨-中山大学-深度学习 2017-07-27 20:36:09**

但是有区别，perceptual loss是针对两张图片

**郑华滨-中山大学-深度学习 2017-07-27 20:36:15**

这里针对的是两组图片

**hinse-scnu-nlp 2017-07-27 20:36:28**

对于未标注真实样本，判别器要给出第k+1个概率低，由于D是用了unlabel的数据。 G为了骗过D, 必须生成和unlabel也比较接近的数据。 因此， 中间层的feature 要接近。所以就显示的增加了feature matching

**hinse-scnu-nlp 2017-07-27 20:36:43**

作为loss

**郑华滨-中山大学-深度学习 2017-07-27 20:36:46**

前者针对两个样本，后者针对的是两个个batch的样本所代表的分布

**Ailsa-多伦多大学-生成模型 2017-07-27 20:37:55**

[强]

**侯瑞兵-计算所-计算机视觉 2017-07-27 20:38:09**

在其他数据集上也会提高分类精度吗

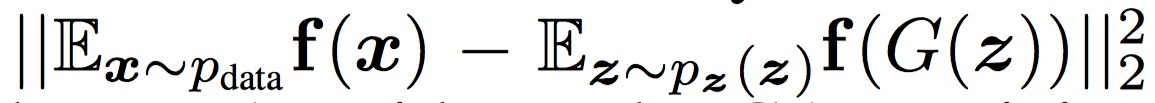
**郑华滨-中山大学-深度学习 2017-07-27 20:38:15**

那其实假feature分布是在追着真feature分布跑对吧，然后真feature分布是想要逃离的

**hinse-scnu-nlp 2017-07-27 20:39:17**

"想要逃离是什么意思？  
"

**郑华滨-中山大学-深度学习 2017-07-27 20:39:37**



**Mashiro-南航-CV 2017-07-27 20:39:37**

与假的有所区分吧

**郑华滨-中山大学-深度学习 2017-07-27 20:39:48**

逃离是指上面这个loss变大

**郑华滨-中山大学-深度学习 2017-07-27 20:39:55**

拉开真假feature的差距

**Ailsa-多伦多大学-生成模型 2017-07-27 20:40:50**

就是判别器尽可能分辨出区别吧

**超-UConn-GAN 2017-07-27 20:41:30**

那半监督的用没有用到NLP和graph上的呢

**郑华滨-中山大学-深度学习 2017-07-27 20:41:33**

嗯，然后在这个过程中如果假feature把有标注真、无标注真两堆feature挤压到一起

**郑华滨-中山大学-深度学习 2017-07-27 20:41:51**

就会促使有标注真、无标注真二者也进行matching

**hinse-scnu-nlp 2017-07-27 20:42:10**

"g是想要减少差距。 d不feature matching. 其实就是 g作弊， 偷看了d的中间曾。 我看看你是怎么分辨我的。 我就生成假的时候，就生成尽量和你D中间层feature接近的。   
"

**郑华滨-中山大学-深度学习 2017-07-27 20:42:18**

半监督文本分类有不少case

**郑华滨-中山大学-深度学习 2017-07-27 20:42:36**

随便搜一下semi supervised text classification就有

**郑华滨-中山大学-深度学习 2017-07-27 20:43:10**

对，但是仅仅这样说还是没法解释怎么帮助ssl的

**郑华滨-中山大学-深度学习 2017-07-27 20:43:58**

我觉得最终得落到有标注真、无标注真两类样本的关系上去解释

**兔子-不存在-ML 2017-07-27 20:44:02**

问一下，那训练好后，是拿D去做分类器来使用吗

**hinse-scnu-nlp 2017-07-27 20:44:02**

"只能理解车成G造假后， D只能更卖力的区分。 把中间层训练的更好  
"

**Gapeng-北京大学-CV 2017-07-27 20:44:04**

作者说这是一个future work，然后就没有下文了[捂脸]

**郑华滨-中山大学-深度学习 2017-07-27 20:44:28**

D直接拿来分类，最后不管第K+1个概率就行

**郑华滨-中山大学-深度学习 2017-07-27 20:44:49**

你这种解释就是论文提的，“为了区分真假而被迫学到的feature，对于分类也会有用”

**郑华滨-中山大学-深度学习 2017-07-27 20:45:19**

但是总感觉不到位[捂脸]

**Gapeng-北京大学-CV 2017-07-27 20:45:33**

毕竟分类器就是D，会有点作用

**hinse-scnu-nlp 2017-07-27 20:45:37**

估计这个问额要遗留了。 只能感觉是G造假能力强了，D也变得厉害了，更接近纳什均衡了。 所以中间层的特征更利于分类。

**hinse-scnu-nlp 2017-07-27 20:46:04**

恩 确实没有很好的证据证明。

**郑华滨-中山大学-深度学习 2017-07-27 20:46:04**

看起来今晚讨论不出了[捂脸]

**Gapeng-北京大学-CV 2017-07-27 20:46:13**

那下一篇吧

**Gapeng-北京大学-CV 2017-07-27 20:47:00**

第一篇有没有同学来讲讲怎么做的？

**Gapeng-北京大学-CV 2017-07-27 20:47:32**

M1和M2各是怎么做的？

**bearbee 清华大学 GAN硬件架构 2017-07-27 20:48:30**

能不能这样解释1.D如果只看输入真实图像和生成图像，太单调，能力不够，加一个feature的损失可能会提高其判别能力。2.G很容易钻空，单纯的学到了一个D分别不出来的图像，就只生成这个，加个feature的话，G要学的东西变多了。

**郑华滨-中山大学-深度学习 2017-07-27 20:49:51**

粗粗扫了一眼，M1是建立样本x到高斯分布上的z的映射，M2是对高斯分布上的z做分类，这样理解对吗

**Ailsa-多伦多大学-生成模型 2017-07-27 20:52:11**

那m2是不是fit MMG呢

**Gapeng-北京大学-CV 2017-07-27 20:52:17**

是不是反过来了？M1是高斯分布z到样本x的映射

**Gapeng-北京大学-CV 2017-07-27 20:52:41**

MMG是什么？

**Ailsa-多伦多大学-生成模型 2017-07-27 20:53:06**

mixture of Gaussian

**郑华滨-中山大学-深度学习 2017-07-27 20:53:16**

哦哦

**郑华滨-中山大学-深度学习 2017-07-27 20:54:19**

诶，能不能简单粗暴地认为，M1是用VAE在有标注、无标注样本上学feature representation，然后M2就在上面直接用有标注样本训练一个分类器，然后它就自动泛化到无标注样本了？

**hinse-scnu-nlp 2017-07-27 20:55:34**

"从3看 ， M1 是x 到 z，   
"

**郑华滨-中山大学-深度学习 2017-07-27 20:55:49**

我论文还没看仔细，先这样猜测一下[捂脸]

**hinse-scnu-nlp 2017-07-27 20:56:15**

然后 （5）就是vae的方法， 两个目标：

**hinse-scnu-nlp 2017-07-27 20:57:15**

1. x生成的z 是normal 或者其他分布（我们定） 2. z生成的和真实样本像

**hinse-scnu-nlp 2017-07-27 20:57:22**

所以就是vae的方法

**hinse-scnu-nlp 2017-07-27 20:58:47**

6的话就是对5在有label情况下的展开， 把y和z一起 生成x

**郑华滨-中山大学-深度学习 2017-07-27 20:59:03**

然后（6）和（7）就是分别针对有标注和无标注样本也学一个变种vae对吗

**郑华滨-中山大学-深度学习 2017-07-27 20:59:14**

学两个vae，说错

**hinse-scnu-nlp 2017-07-27 20:59:17**

(7)对没有label的情况，

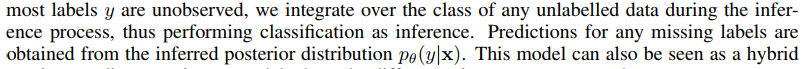
**hinse-scnu-nlp 2017-07-27 20:59:17**

恩

**IIT-lcl-ML 2017-07-27 20:59:17**

小白弱弱的问一下，之前说的D加一个feature的损失可提高判别能力是什么原理？

**Gapeng-北京大学-CV 2017-07-27 20:59:28**



**IIT-lcl-ML 2017-07-27 20:59:41**

可以说是泛化么？

**郑华滨-中山大学-深度学习 2017-07-27 21:00:37**

feature matching的解释后面看总结吧，但其实也没讨论出结果，只有可能的猜想，现在先跳过，讨论第一篇

**Gapeng-北京大学-CV 2017-07-27 21:00:37**

D不加feature的损失，G才加feature matching loss

**hinse-scnu-nlp 2017-07-27 21:00:46**

"D没有加， 是g看了d的中间层 然后g的loss 加了让生成的尽量靠近真实的  
"

**hinse-scnu-nlp 2017-07-27 21:00:58**

恩 是的 一起看看第一篇

**hinse-scnu-nlp 2017-07-27 21:01:40**

（7）没太看懂

**IIT-lcl-ML 2017-07-27 21:02:42**

好好，谢谢

**hinse-scnu-nlp 2017-07-27 21:04:13**

哦 知道了 （7）其实就是把y 当成z的一部分了

**Gapeng-北京大学-CV 2017-07-27 21:04:35**

其实也可以从6直接推出来

**Gapeng-北京大学-CV 2017-07-27 21:05:03**

哦，不行，看错了符号

**hinse-scnu-nlp 2017-07-27 21:05:55**

"（6）有一项 是z|x,y 7是 y,z | x "

**hinse-scnu-nlp 2017-07-27 21:06:58**

就是算KL divergence 时候， 如果有标签， 就把y 和x一起 -> z, 如果没有标签就是 x->y,z

**Gapeng-北京大学-CV 2017-07-27 21:07:14**

"本来想把p\_\theta(x,y)拆成p\_\theta(x)\*p\_\theta(y|x)的，发现似乎导不出来7式"

**Gapeng-北京大学-CV 2017-07-27 21:08:10**

嗯嗯

**hinse-scnu-nlp 2017-07-27 21:08:13**

恩 其实是6,7思路上差异， 有标签，就把y和x一起 生成z。 如果没，就 x生成 y和z

**hinse-scnu-nlp 2017-07-27 21:08:28**

就是把y当成隐含因素了，其实也是合理的。

**hinse-scnu-nlp 2017-07-27 21:09:15**

那对没标签的数据，最后就是看这个隐含变量y了。

**hinse-scnu-nlp 2017-07-27 21:09:27**

是不是啊。 我也看得糊里糊涂的

**Gapeng-北京大学-CV 2017-07-27 21:09:40**

是的

**郑华滨-中山大学-深度学习 2017-07-27 21:09:44**

我目前也只理解到这一层

**Gapeng-北京大学-CV 2017-07-27 21:10:12**

"前面我截了一个图，没有label的数据需要用到p(y|x)"

**Gapeng-北京大学-CV 2017-07-27 21:11:39**

那我们来整理一下思路吧，M1和M2是怎么stack的？整个训练过程是怎样的？

**hinse-scnu-nlp 2017-07-27 21:12:03**

其实是VAE变种， 改变的是 x -> z 要生成norm分布时候。 分了label和unlable, 把x,y 生成 z , 有label就 x->y, z 。 同时要求通过z生成样本尽量真来训练模型。 最后训练完成后得到的latent y 就是无标签的分类了

**hinse-scnu-nlp 2017-07-27 21:12:43**

"中间错了，无label是 x->y, z  
"

**hinse-scnu-nlp 2017-07-27 21:13:32**

peng神整理带大家整理一下， 比较乱

**Gapeng-北京大学-CV 2017-07-27 21:13:49**

。。。。。。

**hinse-scnu-nlp 2017-07-27 21:14:21**

^\_^ 其实为什么叫M1 和 M2 不太理解

**Gapeng-北京大学-CV 2017-07-27 21:14:54**

只是个叫法吧，model 1，model2?

**hinse-scnu-nlp 2017-07-27 21:15:19**

是不是M1 就是有label的x,y->z , M2就是x->y,z ?

**hinse-scnu-nlp 2017-07-27 21:15:46**

"为什么叫discriminative呢？   
"

**郑华滨-中山大学-深度学习 2017-07-27 21:16:05**

我总觉得是不是可以看成一个feature learning模块和一个直接训练分类的模块，因为其实丢掉z，M2的两个loss函数还是可以写的？

**郑华滨-中山大学-深度学习 2017-07-27 21:16:27**

z就是x的一个替身，z跟x放一起并没有增加多余的信息

**郑华滨-中山大学-深度学习 2017-07-27 21:17:16**

所以实际上单独训练M2，形式上也是完整的？只是效果会不好而已？

**Gapeng-北京大学-CV 2017-07-27 21:17:38**

是的

**Gapeng-北京大学-CV 2017-07-27 21:17:47**

直接用M2是可以的

**hinse-scnu-nlp 2017-07-27 21:18:32**

"恩  
"

**郑华滨-中山大学-深度学习 2017-07-27 21:18:38**

此时再看M1加进来的好处，是它跟M2共享了p跟q，共享参数训练不同任务，起到正则化作用

**木羊同学-GDC-SL 2017-07-27 21:18:52**

效果差在什么方面

**郑华滨-中山大学-深度学习 2017-07-27 21:18:52**

不知道这样理解会不会太粗暴[捂脸]

**郑华滨-中山大学-深度学习 2017-07-27 21:19:17**

效果差在没有M1训练任务的正则化约束，会过拟合？

**超-UConn-GAN 2017-07-27 21:19:20**

为什么起到正则化作用呢 不太明白

**郑华滨-中山大学-深度学习 2017-07-27 21:20:03**

我也说不好，就感觉你拿原来任务的一部分参数去训练另一个任务，multi task learning，就会起到正则化作用

**超-UConn-GAN 2017-07-27 21:20:53**

恩 我也感觉像是Multitask的意思

**郑华滨-中山大学-深度学习 2017-07-27 21:21:08**

嗯嗯，要不进入下一篇？

**郑华滨-中山大学-深度学习 2017-07-27 21:21:17**

下一篇感觉好理解一些

**hinse-scnu-nlp 2017-07-27 21:21:31**

"我在想M2怎么用标签  
"

**Gapeng-北京大学-CV 2017-07-27 21:21:31**

还有精力进入下一篇吗？😂

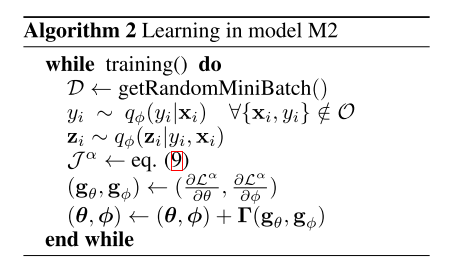
**郑华滨-中山大学-深度学习 2017-07-27 21:21:50**

下一篇真得好理解得多[捂脸]

**兔子-不存在-ML 2017-07-27 21:22:13**

比较难跟上

**Gapeng-北京大学-CV 2017-07-27 21:22:35**



**Gapeng-北京大学-CV 2017-07-27 21:22:48**

先放个M2的训练过程上来，一起探讨一下

**郑华滨-中山大学-深度学习 2017-07-27 21:22:58**

我也只是在强行猜测它的设计思想[捂脸]

**hinse-scnu-nlp 2017-07-27 21:23:00**

我还在想第一份 [Sob]

**hinse-scnu-nlp 2017-07-27 21:23:02**

嗯嗯

**hinse-scnu-nlp 2017-07-27 21:25:30**

感觉M2没用标签， 一定和M1共享参数才行呢？

**Gapeng-北京大学-CV 2017-07-27 21:26:30**

"我的理解是，对于有label的数据，直接跳过y ~ q\_\phi(y|x)，不用算后验了"

**hinse-scnu-nlp 2017-07-27 21:26:55**

恩 是的

**hinse-scnu-nlp 2017-07-27 21:27:24**

"没有标签就要 x先推出y 。   
"

**hinse-scnu-nlp 2017-07-27 21:27:47**

在一起出z。

**Gapeng-北京大学-CV 2017-07-27 21:27:53**

嗯，应该共享参数

**hinse-scnu-nlp 2017-07-27 21:28:23**

我还没看后面的实验。 M1 和 M2 不是独立的？

**Gapeng-北京大学-CV 2017-07-27 21:28:46**

可以独立，也可以stack

**hinse-scnu-nlp 2017-07-27 21:29:46**

"独立用M2， 那有标签的数据不是用不上了。   
"

**hinse-scnu-nlp 2017-07-27 21:30:00**

哦 m2是无监督

**hinse-scnu-nlp 2017-07-27 21:30:55**

"M2只要告诉他类别数， 让他自己学y ？  
"

**郑华滨-中山大学-深度学习 2017-07-27 21:31:03**

如果把z当成x的替身，M2好像就可以独立训练

**郑华滨-中山大学-深度学习 2017-07-27 21:31:17**

M2有同时用到有标注也无标注数据

**hinse-scnu-nlp 2017-07-27 21:31:42**

呵呵 还是没明白 。

**hinse-scnu-nlp 2017-07-27 21:31:52**

M2 怎么使用有标注的数据呢

**郑华滨-中山大学-深度学习 2017-07-27 21:32:23**

用到了y

**郑华滨-中山大学-深度学习 2017-07-27 21:32:26**

公式6就用到了呀

**hinse-scnu-nlp 2017-07-27 21:32:38**

"y i ∼ q φ (y i |x i ) ， 最后校验y？"

**hinse-scnu-nlp 2017-07-27 21:33:07**

"(6)不是M1吗  
"

**Gapeng-北京大学-CV 2017-07-27 21:33:35**

有标注的话，直接跳到这一步

**Gapeng-北京大学-CV 2017-07-27 21:33:35**



**hinse-scnu-nlp 2017-07-27 21:33:40**

哦

**hinse-scnu-nlp 2017-07-27 21:33:52**

"明白了  
"

**郑华滨-中山大学-深度学习 2017-07-27 21:33:55**

公式（5）M1，公式（6）（7）M2

**Gapeng-北京大学-CV 2017-07-27 21:35:19**

这篇比较烧脑[捂脸]

**hinse-scnu-nlp 2017-07-27 21:35:22**

"明白peng神意思了。 有y就省这一步。 无标签的通过共享参数也能学 x->y  
"

**郑华滨-中山大学-深度学习 2017-07-27 21:35:35**

（第三篇跟improved gan那篇完全一样，所以其实就只剩第二篇需要讨论了）

**Gapeng-北京大学-CV 2017-07-27 21:35:40**

讨论以后理解深了好多

**hinse-scnu-nlp 2017-07-27 21:35:40**

"y i ∼ q φ (y i |x i ) ∀{x i , y i } ∈  
/ O  
"

**hinse-scnu-nlp 2017-07-27 21:36:11**

"嗯嗯 是的  
"

**Gapeng-北京大学-CV 2017-07-27 21:36:42**

哦，不，无标签x -> y并不共享参数

**hinse-scnu-nlp 2017-07-27 21:36:44**

"第二篇难不难的？  
"

**hinse-scnu-nlp 2017-07-27 21:36:59**

"恩 我理解 其他共享  
"

**Gapeng-北京大学-CV 2017-07-27 21:37:08**



**郑华滨-中山大学-深度学习 2017-07-27 21:37:11**

比第一篇容易多了哈哈

**hinse-scnu-nlp 2017-07-27 21:37:42**

"x->y 是他独有的。 但是其他参数共享, 让x->y 的分布和有标签的慢慢接近  
"

**郑华滨-中山大学-深度学习 2017-07-27 21:37:55**

感觉（5）和（7）都可以认为是共享参数起正则化作用

**Gapeng-北京大学-CV 2017-07-27 21:37:59**

留到下次吧，后面还有至少4篇，ssl应该要分3次讨论

**郑华滨-中山大学-深度学习 2017-07-27 21:38:17**

咦，还有四篇啊

**hinse-scnu-nlp 2017-07-27 21:38:34**

"我开始看 以为 (5)包括了(6)和(7)  
"

**Gapeng-北京大学-CV 2017-07-27 21:38:34**

嗯，而且协同训练学到的特征应该更有效

**hinse-scnu-nlp 2017-07-27 21:38:39**

回去在慢慢看看

**Gapeng-北京大学-CV 2017-07-27 21:38:52**

之前你们推荐了8篇ssl的文章呀[奸笑]

**hinse-scnu-nlp 2017-07-27 21:39:04**

"一天4篇太狠了  
"

**郑华滨-中山大学-深度学习 2017-07-27 21:39:20**

哦哦刚刚看上面以为只有4篇

**Gapeng-北京大学-CV 2017-07-27 21:39:22**

@hinse-scnu-nlp 要不你来写个总结吧，加深一下理解[奸笑]

**郑华滨-中山大学-深度学习 2017-07-27 21:39:33**

主要第一篇太难了[捂脸]

**Gapeng-北京大学-CV 2017-07-27 21:39:44**

是啊

**hinse-scnu-nlp 2017-07-27 21:40:02**

其实还有个问题

**郑华滨-中山大学-深度学习 2017-07-27 21:40:02**

感觉碰到这种变分推断的东西就很容易被绕晕

**hinse-scnu-nlp 2017-07-27 21:40:11**

"Latent Feature Discriminative Model Objective  
  
Generative Semi-supervised Model Objective"

**hinse-scnu-nlp 2017-07-27 21:40:17**

名字很绕

**hinse-scnu-nlp 2017-07-27 21:40:28**

一个叫D一个叫G

**郑华滨-中山大学-深度学习 2017-07-27 21:40:40**

我觉得就如其所言，M1就是用来学feature的

**Gapeng-北京大学-CV 2017-07-27 21:40:45**

哈哈

**郑华滨-中山大学-深度学习 2017-07-27 21:40:57**

然后顺便共享参数正则化一下

**hinse-scnu-nlp 2017-07-27 21:41:33**

"哈哈 还好你们带着看。 确实好难，谁推荐的。   
"

**hinse-scnu-nlp 2017-07-27 21:41:44**

要赞一个。

**郑华滨-中山大学-深度学习 2017-07-27 21:41:54**

（谁推荐的站出来）

**hinse-scnu-nlp 2017-07-27 21:42:29**

其实开始感觉是 VAE的稍微变化， 但越看越迷糊。 我先回家了

**Gapeng-北京大学-CV 2017-07-27 21:43:03**

@陆鹏起-hust-gan 你推荐的[奸笑]

**Gapeng-北京大学-CV 2017-07-27 21:43:32**

回家写总结[奸笑]

**Gapeng-北京大学-CV 2017-07-27 21:44:02**

@hinse-scnu-nlp

**Gapeng-北京大学-CV 2017-07-27 21:45:24**

@hinse-scnu-nlp 你跟@郑华滨-中山大学-深度学习 是好搭档呀，之前又一次讨论也是[奸笑]