Semi-supervised learning(part 3) and SCAN.

**Gapeng-北京大学-CV 2017-08-17 20:00:33**



**Gapeng-北京大学-CV 2017-08-17 20:00:43**

搞事情！搞事情！

**starif-西电-ml 2017-08-17 20:00:51**

如果毛豆大佬还在线下吸猫..

**Gapeng-北京大学-CV 2017-08-17 20:01:18**

大家好，又到了8点了，我们开始SSL第三次讨论吧。

**Gapeng-北京大学-CV 2017-08-17 20:01:24**

今天的主题不只SSL，还有一篇非监督学习的文章SCAN，上次大家推荐的。

**Gapeng-北京大学-CV 2017-08-17 20:01:47**

今天SSL讨论的文章是Good Semi-supervised Learning That Requires a Bad GAN(1705.09783)。我们先从这篇开始。这篇是毛豆大佬的最爱

**hwang-中南大学-生成模型 2017-08-17 20:02:29**

毛豆大佬呢，出来搞事情辣

**hinse-scnu-nlp 2017-08-17 20:02:42**

为啥子最爱啊

**Hao 佐治亚大学 量子计算 2017-08-17 20:02:47**

搬了小板凳围观学习

**兔子-不存在-ML 2017-08-17 20:02:50**

毛豆出来开头呀

**Gapeng-北京大学-CV 2017-08-17 20:02:58**

首先，标题是啥意思？

**anshiquanshu66-师大-医学图像处理 2017-08-17 20:03:21**

搬小板凳围观学习

**兔子-不存在-ML 2017-08-17 20:03:43**

其实只是需要坏的G，D还是好好的

**hinse-scnu-nlp 2017-08-17 20:03:51**

bad gan有助于ssl

**hwang-中南大学-生成模型 2017-08-17 20:03:57**

好白菜需要猪来拱

**Hao 佐治亚大学 量子计算 2017-08-17 20:04:05**

perfect generator无法有更好的generalization

**Gapeng-北京大学-CV 2017-08-17 20:04:08**

哈哈哈

**兔子-不存在-ML 2017-08-17 20:04:20**

然后坏的标准是G的分布不匹配真实分布

**hinse-scnu-nlp 2017-08-17 20:04:53**

刻意的避开 结论挺意外的

**陆鹏起-hust-gan 2017-08-17 20:05:04**

准确的说是improved-gan的目标函数优化不了更好的generalization error

**Gapeng-北京大学-CV 2017-08-17 20:05:14**

嗯呐，这是文章的第一个定理证明的，GAN的生成学太好了，相当于只有监督学习的部分起作用

**兔子-不存在-ML 2017-08-17 20:05:49**

这个有没有示例代码

**starif-西电-ml 2017-08-17 20:05:59**

这里的bad 指的是不学源数据分布, 而是要学到它的补集，用毛豆大佬的话说 叫做隔离带

**Gapeng-北京大学-CV 2017-08-17 20:06:13**

我觉得他说的不是improved GAN优化不了

**Gapeng-北京大学-CV 2017-08-17 20:06:26**

而是improved gan

**木羊同学-GDC-SL 2017-08-17 20:06:32**

Bad有个明确定义的

**Gapeng-北京大学-CV 2017-08-17 20:06:50**

而是improved gan的目标函数可能存在问题

**兔子-不存在-ML 2017-08-17 20:07:08**

这个补集怎么求出来

**Gapeng-北京大学-CV 2017-08-17 20:07:11**

再mnist上问题似乎不大，也就是improved gan那篇做的实验上问题不大

**Gapeng-北京大学-CV 2017-08-17 20:07:42**

这个补集直接算是求不了的

**人5人6-CMCC-nlp 2017-08-17 20:07:44**

G完美，未标记数据不能有效被利用

**hwang-中南大学-生成模型 2017-08-17 20:07:53**

论文也没仔细看，说点个人理解，欢迎拍砖：直观上来看，既然最终目标是做分类，那就要把重点放在分类的准确率上。一般来说，在训练数据中出现的较少的那些sample上，分类器由于更新次数不足，表现得往往比较差通过互补分布的思想，把真实分布中较少出现的数据变多，加强分类器在这些sample上的更新，从而可以提升分类器在这些sample上的表现，提升整体的分类正确率。

**Gapeng-北京大学-CV 2017-08-17 20:08:13**

我们先来明确一下，bad gan是什么含义？

**hinse-scnu-nlp 2017-08-17 20:08:47**

g不匹配真实分布

**陆鹏起-hust-gan 2017-08-17 20:08:58**

bad generalization error

**Gapeng-北京大学-CV 2017-08-17 20:09:04**

说具体一点

**starif-西电-ml 2017-08-17 20:09:19**

bad gan 的对立面是 Perfect Generator

**hinse-scnu-nlp 2017-08-17 20:09:29**

隔离带

**hinse-scnu-nlp 2017-08-17 20:10:18**

一个个孤岛之间的湖水？

**Gapeng-北京大学-CV 2017-08-17 20:10:49**

可以，这个理解很形象

**兔子-不存在-ML 2017-08-17 20:11:18**

还是没理解……

**Gapeng-北京大学-CV 2017-08-17 20:11:40**

不着急，我们一步步来

**Hao 佐治亚大学 量子计算 2017-08-17 20:11:45**

和perfect generator互补的generator

**Gapeng-北京大学-CV 2017-08-17 20:12:01**

讲到目标函数的时候就能理解了

**Gapeng-北京大学-CV 2017-08-17 20:12:04**

为了学一个bad gan，文章用的是complement generator，这个又是怎么定义的？

**hinse-scnu-nlp 2017-08-17 20:12:09**

figure2 左边的图

**庄胤-北京理工大学-遥感处理 2017-08-17 20:12:10**

文中说分布一致的样本对discrimination没有任何提升不是很理解，难道要分布不一致才能提升？

**Hao 佐治亚大学 量子计算 2017-08-17 20:12:10**

一个个岛屿就是一个个已知的类

**starif-西电-ml 2017-08-17 20:12:57**

里面有个假定 各个类别之间的数据是隔离开的

**hinse-scnu-nlp 2017-08-17 20:12:58**

想想是有道理的 且听群主慢慢引导

**Gapeng-北京大学-CV 2017-08-17 20:13:07**

分布一致的话，直观上来说，相当于就跟原来的数据一样

**hinse-scnu-nlp 2017-08-17 20:14:13**

让各个分类尽可能的确定 g就分布在各个分类的间隔中

**丁铭 清华 数据挖掘 2017-08-17 20:15:32**

疑问，补集分布中归一化因子z被忽略了，这会影响熵和另一部分的相对大小吧？

**Hao 佐治亚大学 量子计算 2017-08-17 20:16:14**

complement生成器定义怎么理解啊

**hinse-scnu-nlp 2017-08-17 20:16:24**

只是为了概率和为1吧

**Gapeng-北京大学-CV 2017-08-17 20:16:52**

不会影响

**丁铭 清华 数据挖掘 2017-08-17 20:17:15**

为什么

**Gapeng-北京大学-CV 2017-08-17 20:18:25**

分布确定了，Z就确定了

**Gapeng-北京大学-CV 2017-08-17 20:18:31**

是一个常数

**hinse-scnu-nlp 2017-08-17 20:18:33**

没忽略吧

**hinse-scnu-nlp 2017-08-17 20:19:23**

公式4中平均时候z就没了

**丁铭 清华 数据挖掘 2017-08-17 20:19:23**

但是z并不能求出来，他是确定的又能如何

**hinse-scnu-nlp 2017-08-17 20:19:37**

因为他是常数

**丁铭 清华 数据挖掘 2017-08-17 20:20:29**

但是p是要优化的吧

**Gapeng-北京大学-CV 2017-08-17 20:21:06**

p是固定的，p\_G才是要优化的

**hinse-scnu-nlp 2017-08-17 20:21:26**

p是真实分布

**Gapeng-北京大学-CV 2017-08-17 20:21:34**

p是真实样本的密度函数

**丁铭 清华 数据挖掘 2017-08-17 20:21:38**

说错了，我就说pg啊

**丁铭 清华 数据挖掘 2017-08-17 20:22:12**

ok，懂了

**hinse-scnu-nlp 2017-08-17 20:22:38**

没看那么细 z和c我看时候都略过了

**丁铭 清华 数据挖掘 2017-08-17 20:22:38**

忘记概率加起来等于1了[捂脸]

**Gapeng-北京大学-CV 2017-08-17 20:22:39**

好，我们继续

**hinse-scnu-nlp 2017-08-17 20:23:01**

嗯

**Gapeng-北京大学-CV 2017-08-17 20:24:11**

按照原来improved gan的那一套做法，可能会出现什么问题？

**Gapeng-北京大学-CV 2017-08-17 20:25:17**

我们知道，improved gan里面提到，如果不加feature matching(FM)的话，会很差

**Gapeng-北京大学-CV 2017-08-17 20:25:55**

这篇文章指出，feature matching虽好，但是也存在问题

**Gapeng-北京大学-CV 2017-08-17 20:26:03**

那么究竟是什么问题？

**hinse-scnu-nlp 2017-08-17 20:27:06**

其实没太明白 最终他也用了fm

**Gapeng-北京大学-CV 2017-08-17 20:27:43**

嗯，fm还是要用的

**hinse-scnu-nlp 2017-08-17 20:28:36**

我是这么理解的 像图2

**hinse-scnu-nlp 2017-08-17 20:29:10**

fm让g的颜色和真实分布的颜色接近

**hinse-scnu-nlp 2017-08-17 20:29:44**

但还是和真实分布还是要留出边界

**Gapeng-北京大学-CV 2017-08-17 20:29:58**

对，但是太接近的时候，D就会把真实数据分成fake

**hinse-scnu-nlp 2017-08-17 20:30:23**

g生成的面积大 因此概率密度低于真实类别的概率密度

**Gapeng-北京大学-CV 2017-08-17 20:31:01**

另一个问题，G的生成不能避免mode collapse

**hinse-scnu-nlp 2017-08-17 20:31:16**

但原来问题在哪还要老大道破

**Gapeng-北京大学-CV 2017-08-17 20:31:53**

你说不加fm的问题吗？

**hinse-scnu-nlp 2017-08-17 20:32:25**

不是 其实你说答案了

**hinse-scnu-nlp 2017-08-17 20:32:35**

太接近时候...

**hwang-中南大学-生成模型 2017-08-17 20:32:50**

受教

**Gapeng-北京大学-CV 2017-08-17 20:33:01**

这篇文章就是针对这两个问题做了改进

**hinse-scnu-nlp 2017-08-17 20:33:33**

嗯 增加了g的不确定度作为目标

**Gapeng-北京大学-CV 2017-08-17 20:33:36**

我们来看一下他是解决这个问题的

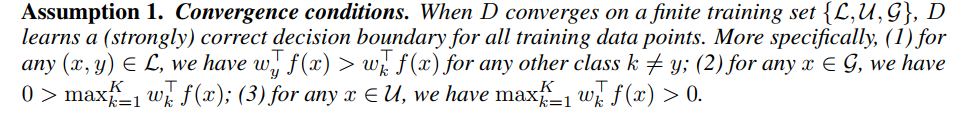
**hinse-scnu-nlp 2017-08-17 20:34:19**

我做的高铁好晃有点头晕

**hinse-scnu-nlp 2017-08-17 20:34:33**

pt

**Gapeng-北京大学-CV 2017-08-17 20:34:50**



**hinse-scnu-nlp 2017-08-17 20:34:52**

pullaway

**Gapeng-北京大学-CV 2017-08-17 20:35:09**

这是文章要达到的目标

**Gapeng-北京大学-CV 2017-08-17 20:35:23**

@hinse-scnu-nlp PT loss[奸笑]

**Gapeng-北京大学-CV 2017-08-17 20:36:03**

同时还要避免FM的那两个问题

**Gapeng-北京大学-CV 2017-08-17 20:36:39**

因此作者说要学一个bad gan

**hinse-scnu-nlp 2017-08-17 20:36:45**

嗯 嗯

**Gapeng-北京大学-CV 2017-08-17 20:37:33**

Bad gan的分布跟真实分布不能对齐，对齐匹配上了就变成perfect gan了

**Gapeng-北京大学-CV 2017-08-17 20:38:07**

这个互补的gan是通过什么目标函数来实现的？

**丁铭 清华 数据挖掘 2017-08-17 20:38:15**

改进feature match这跟pt loss有什么关系哇，pt不是来近似求熵的嘛

**丁铭 清华 数据挖掘 2017-08-17 20:38:24**

求大佬再讲得详细点

**hinse-scnu-nlp 2017-08-17 20:38:27**

感觉已经不太像gan了

**hinse-scnu-nlp 2017-08-17 20:39:09**

pt是为了防collapse

**hinse-scnu-nlp 2017-08-17 20:39:32**

增加生成样本的不确定性

**丁铭 清华 数据挖掘 2017-08-17 20:39:53**

原本是这样的哇

**丁铭 清华 数据挖掘 2017-08-17 20:40:13**

可是我好像看到这篇文章中不是这么说的QAQ

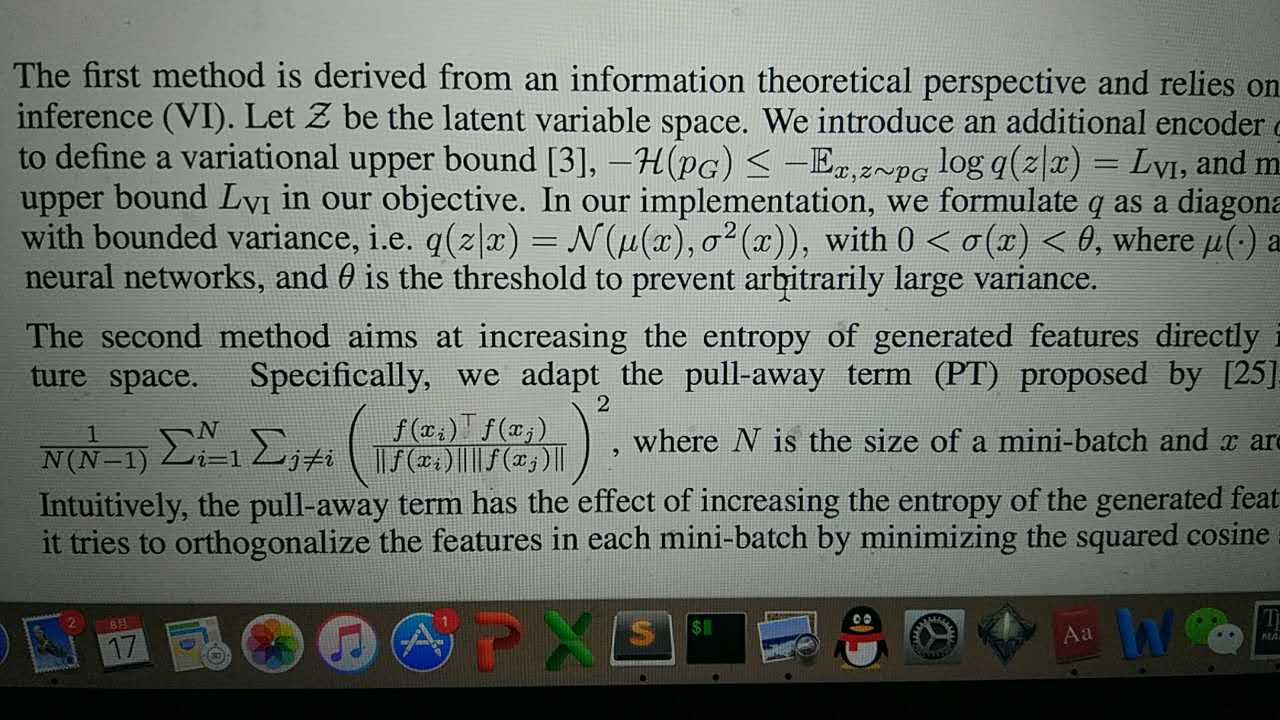
**Gapeng-北京大学-CV 2017-08-17 20:40:15**

pt：pull away，就是让生成样本尽可能地差异大

**hinse-scnu-nlp 2017-08-17 20:40:39**

先下高铁

**丁铭 清华 数据挖掘 2017-08-17 20:41:27**



**丁铭 清华 数据挖掘 2017-08-17 20:42:04**

他的意思是，我loss里边有一项是熵，但是很难估计，现在我直接用pt近似，反正效果差不多，是这样吧？

**hinse-scnu-nlp 2017-08-17 20:43:14**

嗯 里面很多近似

**Gapeng-北京大学-CV 2017-08-17 20:43:33**

嗯，pt是间接地最大化熵

**hinse-scnu-nlp 2017-08-17 20:44:15**

嗯嗯

**Gapeng-北京大学-CV 2017-08-17 20:45:08**

优化pt是沿着熵增地方向在走

**Gapeng-北京大学-CV 2017-08-17 20:45:55**

我有点引导不下去了

**Gapeng-北京大学-CV 2017-08-17 20:46:01**

你们提问题吧

**Gapeng-北京大学-CV 2017-08-17 20:46:03**

一起讨论

**黄瑞阳\_郑州大学\_NLP 2017-08-17 20:46:14**

pull-away term (PT) proposed by [25],

**wrj-西电-UDA 2017-08-17 20:46:19**



**黄瑞阳\_郑州大学\_NLP 2017-08-17 20:46:22**

PT的方法文中没有介绍。

**wrj-西电-UDA 2017-08-17 20:46:24**

B是什么意思？

**Gapeng-北京大学-CV 2017-08-17 20:46:57**

B是凸包

**Gapeng-北京大学-CV 2017-08-17 20:47:43**

PT不是都给了loss项了么，有了loss就能优化了

**丁铭 清华 数据挖掘 2017-08-17 20:48:11**

pt那个刚提出来的时候是直接当成正则项强加的，这篇文章是代替熵是loss。我觉得从出发点上还是不一样的

**wrj-西电-UDA 2017-08-17 20:48:21**

这个凸包里有没有元素 不属于Fk的并？

**hinse-scnu-nlp 2017-08-17 20:48:24**

@黄瑞阳\_郑州大学\_NLP 也算有了 feature空间相似 其实和主题贴近。分割岛的湖水 就比较类似 但是靠近每个岛的水和这个岛要类似就是fm

**丁铭 清华 数据挖掘 2017-08-17 20:48:32**

反正都是说一套做一套嘛😂

**Gapeng-北京大学-CV 2017-08-17 20:48:51**

@wrj-西电-UDA 可以有

**wrj-西电-UDA 2017-08-17 20:49:06**

好的

**丁铭 清华 数据挖掘 2017-08-17 20:49:09**

详细一下一个大圆包含这些孤岛

**丁铭 清华 数据挖掘 2017-08-17 20:49:18**

想像

**丁铭 清华 数据挖掘 2017-08-17 20:49:37**

想象

**丁铭 清华 数据挖掘 2017-08-17 20:50:19**

pixelCNN++谁能介绍一下

**丁铭 清华 数据挖掘 2017-08-17 20:50:30**

我不是搞cv的，确实不太懂

**Gapeng-北京大学-CV 2017-08-17 20:50:32**

这篇我也没看

**hinse-scnu-nlp 2017-08-17 20:51:19**



**Gapeng-北京大学-CV 2017-08-17 20:52:46**

pixelCNN++有没有同学看过的，求介绍一下

**丁铭 清华 数据挖掘 2017-08-17 20:53:26**

另外你们觉得用feature matching的方法把pg拖到B里是好的吗？我觉得不太优雅

**Gapeng-北京大学-CV 2017-08-17 20:53:55**

为什么？

**丁铭 清华 数据挖掘 2017-08-17 20:55:24**

主要是没有什么理论保证吧

**hinse-scnu-nlp 2017-08-17 20:55:24**

fm还是要的 g要生成特征类似于某个真实类别。 不然就没gan的感觉了。只是分布比较稀疏

**hinse-scnu-nlp 2017-08-17 20:55:51**

要给d一点难度

**丁铭 清华 数据挖掘 2017-08-17 20:55:54**

而且某种程度上跟前面两项有点冲突，虽然也不是不行

**Gapeng-北京大学-CV 2017-08-17 20:56:05**

feature matching把生成分布被扯到真实分布附近

**丁铭 清华 数据挖掘 2017-08-17 20:56:44**

我直觉上觉得可以有更优雅的扯法

**hinse-scnu-nlp 2017-08-17 20:56:49**

嗯 我开始也觉得有冲突 但看了图2觉得也并不矛盾

**丁铭 清华 数据挖掘 2017-08-17 20:56:52**

但我得想想

**Gapeng-北京大学-CV 2017-08-17 20:57:09**

[机智]

**丁铭 清华 数据挖掘 2017-08-17 20:57:16**

当然我不是说这么做有问题

**Gapeng-北京大学-CV 2017-08-17 20:57:43**

另外，conditional entropy那个项就是毛豆大佬常说的，要求尖峰分布，任何一个都可以。[奸笑]

**hinse-scnu-nlp 2017-08-17 20:57:46**

嗯 是可能有更好的方法

**丁铭 清华 数据挖掘 2017-08-17 20:57:46**

只是优雅程度不如前边

**Gapeng-北京大学-CV 2017-08-17 20:58:13**

毛豆大佬不在的时候，要时刻惦记着

**丁铭 清华 数据挖掘 2017-08-17 20:58:47**

毛豆是谁

**丁铭 清华 数据挖掘 2017-08-17 20:58:54**

⊙ω⊙

**Gapeng-北京大学-CV 2017-08-17 20:58:54**

是呢，前面的理论上都比较漂亮，这里就像一个黑盒子

**黄瑞阳\_郑州大学\_NLP 2017-08-17 20:58:55**

看了实验部分，精度算目前不错的，是不是半监督中这篇算最高的？

**Gapeng-北京大学-CV 2017-08-17 20:59:08**

@郑华滨-中山大学-深度学习

**hinse-scnu-nlp 2017-08-17 20:59:08**

要d更确信他的分类

**丁铭 清华 数据挖掘 2017-08-17 20:59:10**

肯定得是吧

**丁铭 清华 数据挖掘 2017-08-17 20:59:34**

毕竟kimi出品，必属精品😂

**hinse-scnu-nlp 2017-08-17 20:59:43**

warbean 战斗着的毛豆 可惜今晚不在

**Gapeng-北京大学-CV 2017-08-17 20:59:57**

最近的一篇SSL应该就是它了吧？

**丁铭 清华 数据挖掘 2017-08-17 21:00:09**

噢噢，我好像在知乎看过他的Wgan

**hinse-scnu-nlp 2017-08-17 21:00:09**

这篇是很巧妙 还有引用的3也是

**hinse-scnu-nlp 2017-08-17 21:01:00**

本来今晚看你们讨论的 才发现毛豆不知道哪里逗猫去了

**丁铭 清华 数据挖掘 2017-08-17 21:01:30**

3说的是啥，ebgan为啥有矫正啊

**Gapeng-北京大学-CV 2017-08-17 21:01:36**

一会儿吸猫他可能就出来了

**陆鹏起-hust-gan 2017-08-17 21:01:44**

看到一些以前知乎上关于变分推断的问题，都能看到warbean大佬疯狂点赞一堆回答

**Gapeng-北京大学-CV 2017-08-17 21:03:34**

关于这篇还有问题吗？

**hinse-scnu-nlp 2017-08-17 21:04:17**

我要赶去机场了 回头看大家总结

**hwang-中南大学-生成模型 2017-08-17 21:04:20**

有木有谁跑过这个模型

**kk+国科大+NLP 2017-08-17 21:04:22**

为什么大家聊的好深，感觉小白才进来，跟不上怎么办

**郑华滨-中山大学-深度学习 2017-08-17 21:04:28**

今晚开会去了[捂脸][捂脸][捂脸][捂脸]

**hinse-scnu-nlp 2017-08-17 21:04:55**

晕 我跑了你才过来

**hinse-scnu-nlp 2017-08-17 21:05:26**

毛豆对模型理解很深 快发表一下 我回头看总结

**郑华滨-中山大学-深度学习 2017-08-17 21:05:31**

最近被抓去做人脸都没怎么关注gan了[捂脸]

**Gapeng-北京大学-CV 2017-08-17 21:05:31**

猫豆有啥要说的吗？

**Gapeng-北京大学-CV 2017-08-17 21:05:57**

关于bad gan这篇

**郑华滨-中山大学-深度学习 2017-08-17 21:05:59**

我……你们讨论得很好！我要学习一个！

**Gapeng-北京大学-CV 2017-08-17 21:06:14**

。。。

**丁铭 清华 数据挖掘 2017-08-17 21:06:38**

大佬就讲一下pixelCNN++吧[机智]

**Gapeng-北京大学-CV 2017-08-17 21:06:48**

你之前说要对比这篇和CatGAN，TripleGAN的

**Gapeng-北京大学-CV 2017-08-17 21:06:53**

我都记着

**Guo-Jun Qi 2017-08-17 21:07:14**

"我大概看了这篇论文，还没看完，有个问题：求complement generator和generator难道不是一回事？有了生成互补分布的generator，本质上难道不是排除掉false sample所在的区域后，给出了true distribution?  
"

**kk+国科大+NLP 2017-08-17 21:07:21**

求指教

**郑华滨-中山大学-深度学习 2017-08-17 21:07:27**

我记得上次我的想法是bad gan能够逼判别器但是又不会完全插入进去导致判别器懵逼掉……[捂脸]

**郑华滨-中山大学-深度学习 2017-08-17 21:08:15**

它的互补应该是在一个限定的全集下互补？

**Guo-Jun Qi 2017-08-17 21:08:22**

yes

**Guo-Jun Qi 2017-08-17 21:08:35**

所以引入个convex hull假设

**郑华滨-中山大学-深度学习 2017-08-17 21:08:40**

（还没看上面的讨论，瞎说一下）

**Gapeng-北京大学-CV 2017-08-17 21:08:46**

嗯

**丁铭 清华 数据挖掘 2017-08-17 21:09:05**

啊啊？是啊，没看懂您的问题[捂脸]

**Guo-Jun Qi 2017-08-17 21:09:18**

"但convex hull在算法里没有体现  
"

**Gapeng-北京大学-CV 2017-08-17 21:09:23**

我也没看懂。。。

**丁铭 清华 数据挖掘 2017-08-17 21:09:49**

确实，所以我说那部分不优美

**kk+国科大+NLP 2017-08-17 21:09:54**

我想问一下，是不是ssl是另一个理解的gan？

**丁铭 清华 数据挖掘 2017-08-17 21:10:21**

他时间feature matching到真实分布“附近”

**丁铭 清华 数据挖掘 2017-08-17 21:10:29**

\*直接

**郑华滨-中山大学-深度学习 2017-08-17 21:10:38**

哦哦我还想起来不要逼得太紧就会有一个结果，即g只会往真实数据的稀疏区域进攻，而不会入侵到稠密区域，这样子有利于d的特征空间上进行标签传播

**郑华滨-中山大学-深度学习 2017-08-17 21:11:22**

g在真实数据稀疏区域建立隔离带，但是隔离带又不会延伸到稠密区域，阻碍标签传播

**郑华滨-中山大学-深度学习 2017-08-17 21:11:55**

不知道上面有没有讨论标签传播的角度

**Gapeng-北京大学-CV 2017-08-17 21:13:03**

@Guo-Jun Qi generator的分布跟真实分布有重叠，他这里要求要互补，尽可能少重叠，原来的generator没有这个限制

**泛艺术范儿-西电-DL 2017-08-17 21:13:15**

论文里是把概率支撑集中概率小于某个阈值的部分去掉。然后在求互补。相当于生成器侧重于生成概率支撑集里概率稀疏的部分。

**Gapeng-北京大学-CV 2017-08-17 21:13:22**

不过他也没有限制是凸包

**kk+国科大+NLP 2017-08-17 21:13:43**

互补的概念是怎么样的集合

**Gapeng-北京大学-CV 2017-08-17 21:15:10**

就等你来说呢@郑华滨-中山大学-深度学习 [机智]

**Gapeng-北京大学-CV 2017-08-17 21:15:39**

@kk+国科大+NLP 没懂你的问题

**泛艺术范儿-西电-DL 2017-08-17 21:15:47**



**泛艺术范儿-西电-DL 2017-08-17 21:15:56**



**Gapeng-北京大学-CV 2017-08-17 21:17:06**

还有问题吗？

**kk+国科大+NLP 2017-08-17 21:17:14**

就是补集是算不算上真实数据集的，当G不是很好的时候，

**kk+国科大+NLP 2017-08-17 21:17:42**

G和真实数据空间应该不在同一流形吧

**GMiaow-鹅厂-CV 2017-08-17 21:18:21**

😄

**Guo-Jun Qi 2017-08-17 21:18:56**

"本质上，这个 bad GAN SSL是用生成的false samples（相对于true samples)来作为所有类别(1到K)的负例。但一般来说，负样本可能提供不了太有用的信息，它们主要是用来排除不可能的区域。真正有用的是接近当前分界面的样本，这样样本的分类熵比较大，所有我觉得真正在这个bad gan里起作用的是那个entropy term  
"

**kk+国科大+NLP 2017-08-17 21:20:34**

厉害👍

**Gapeng-北京大学-CV 2017-08-17 21:20:52**

conditional entropy term？

**陆鹏起-hust-gan 2017-08-17 21:21:10**

同感

**Guo-Jun Qi 2017-08-17 21:21:27**

"P\_G的entropy,应该是conditional entropy  
"

**Guo-Jun Qi 2017-08-17 21:21:36**

具体实现细节没看

**丁铭 清华 数据挖掘 2017-08-17 21:21:41**

g的那个吧

**kk+国科大+NLP 2017-08-17 21:21:42**

这一层确实没想到

**丁铭 清华 数据挖掘 2017-08-17 21:21:44**

嗯

**Gapeng-北京大学-CV 2017-08-17 21:21:56**

哦，p\_G那个

**丁铭 清华 数据挖掘 2017-08-17 21:22:11**

那个不是conditional的

**Guo-Jun Qi 2017-08-17 21:22:35**

我觉得方向对的，这个我们组其实现在就在用类似的方法；但分析我觉得不是太对

**郑华滨-中山大学-深度学习 2017-08-17 21:22:38**

完全原理真样本的假样本肯定没啥用

**郑华滨-中山大学-深度学习 2017-08-17 21:23:06**

主要问题是一头撞进真是样本稠密区域的假样本有没有害

**Guo-Jun Qi 2017-08-17 21:23:25**

离当前分界面很远的负样本, or false samples，肯定没啥用

**郑华滨-中山大学-深度学习 2017-08-17 21:23:29**

真是->真实

**丁铭 清华 数据挖掘 2017-08-17 21:24:19**

可是您为什么就觉得生成的样本离分界面远？

**丁铭 清华 数据挖掘 2017-08-17 21:24:39**

是fm的手段不够强？

**Gapeng-北京大学-CV 2017-08-17 21:24:47**

齐老师这个解读很不错

**Guo-Jun Qi 2017-08-17 21:25:32**

"这个方法没显示的分析complement G得到的样本相对分界面的距离关系  
"

**Guo-Jun Qi 2017-08-17 21:26:07**

所有，我既不能说是距离近的负样本，也不能说是距离远的负样本

**Guo-Jun Qi 2017-08-17 21:26:40**

当然，bad gan也没办法说这点

**丁铭 清华 数据挖掘 2017-08-17 21:26:54**

嗯，我一开始说得也是这个问题。直接扯过去太不优雅

**Guo-Jun Qi 2017-08-17 21:27:51**

没有偏向性的论断就是啥都没说

**丁铭 清华 数据挖掘 2017-08-17 21:28:45**

反正目标就是生成孤岛间的空隙嘛😂这个想法我觉得还是很对的

**kk+国科大+NLP 2017-08-17 21:29:33**

各位大佬，理解很深入

**Gapeng-北京大学-CV 2017-08-17 21:30:01**

所以，是不是可以说，生成边界附近的样本就行了

**郑华滨-中山大学-深度学习 2017-08-17 21:30:32**

感觉是要求d的特征在孤岛间突变，在孤岛内渐变

**丁铭 清华 数据挖掘 2017-08-17 21:30:34**

连齐老师都粗来了，我这种菜鸡还是先潜水好了(●°u°●)​ 」

**kk+国科大+NLP 2017-08-17 21:30:45**

刚入门的小白要怎么可以看得这么多角度，或者该怎么看论文，我发现我都没有看出这些来

**郑华滨-中山大学-深度学习 2017-08-17 21:31:06**

然后就完成了标签传播（我还是比较喜欢从这个角度解释[捂脸]）

**Guo-Jun Qi 2017-08-17 21:31:12**

其实bad gan就是做减法， 减去不可能的负样本区域。相对与用good gan做加法，肯定做减法去生成负样本跟容易。但容易是容易，也容易搞出很多的trivial negative samples

**郑华滨-中山大学-深度学习 2017-08-17 21:31:57**

有道理

**Gapeng-北京大学-CV 2017-08-17 21:32:24**

所以他才需要p\_G的entropy term

**丁铭 清华 数据挖掘 2017-08-17 21:32:33**

默默问一句，加法怎么做😂

**丁铭 清华 数据挖掘 2017-08-17 21:32:41**

是哪篇文章

**Guo-Jun Qi 2017-08-17 21:32:45**

比如，最trivial 的negative samples是黑屏。。。

**Gapeng-北京大学-CV 2017-08-17 21:33:01**

哈哈哈

**Guo-Jun Qi 2017-08-17 21:33:09**

"之前的gan based ssl基本都是加法  
"

**郑华滨-中山大学-深度学习 2017-08-17 21:33:11**

😂

**丁铭 清华 数据挖掘 2017-08-17 21:33:40**

啊啊，最基本的不是把fake当成一类吗？

**Gapeng-北京大学-CV 2017-08-17 21:34:15**

从它学习的结果来看，是不是可以认为，negative samples还是在边界附近的？

**Guo-Jun Qi 2017-08-17 21:34:20**

但目标是使得G是生成真实的样本

**Guo-Jun Qi 2017-08-17 21:34:36**

也许是因为entropy term?

**丁铭 清华 数据挖掘 2017-08-17 21:34:47**

如果是good gan，生成真实样本，那么d仍应该分到最后一类

**Guo-Jun Qi 2017-08-17 21:34:55**

我还没全部看完细节，需要点时间研究下

**郑华滨-中山大学-深度学习 2017-08-17 21:34:58**

我最近在看检测，感觉检测领域里面的hard sample mining跟这个bad gan很有联系，它是说每次做二分类训练的时候，只对loss排名靠前的样本进行梯度更新，因为它们是hard example

**丁铭 清华 数据挖掘 2017-08-17 21:34:58**

还是减法吧

**starif-西电-ml 2017-08-17 21:35:48**

加法 减法是相对g目标而言的吧

**丁铭 清华 数据挖掘 2017-08-17 21:37:02**

我始终没明白good gan的理论，训练得好了对分类的帮助，直观上我没有理解(>﹏<)

**Gapeng-北京大学-CV 2017-08-17 21:37:45**

训练太好了，反而没有帮助吧

**Gapeng-北京大学-CV 2017-08-17 21:38:36**

bad gan这篇一开始就证明了，训练太好了，泛化误差跟直接在有标签样本上监督是一样的

**Guo-Jun Qi 2017-08-17 21:39:28**

"我觉得，目标(4)训练G，没有体现complement generator得想法  
"

**Guo-Jun Qi 2017-08-17 21:40:53**

第二项难道不是还是定义true distribution p(x)得support set?

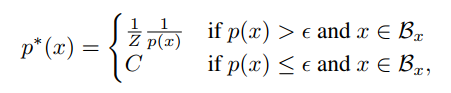
**Guo-Jun Qi 2017-08-17 21:42:13**

不过这部分看得有点快，也可能我理解有误

**丁铭 清华 数据挖掘 2017-08-17 21:42:29**

什么意思

**Gapeng-北京大学-CV 2017-08-17 21:42:33**



**泛艺术范儿-西电-DL 2017-08-17 21:42:49**

他设了个阈值，support set里大于阈值的部分

**Gapeng-北京大学-CV 2017-08-17 21:42:51**

complement体现在这里

**丁铭 清华 数据挖掘 2017-08-17 21:42:58**

第二项就是个交叉熵

**丁铭 清华 数据挖掘 2017-08-17 21:43:17**

互补主要就体现在这一项啊

**丁铭 清华 数据挖掘 2017-08-17 21:43:34**

@Guo-Jun Qi

**Gapeng-北京大学-CV 2017-08-17 21:44:19**

第二项是说，对于密度高的地方，P\_g不能往这边走

**starif-西电-ml 2017-08-17 21:44:24**

第二项不就是尽量避开px的高密度分布区吗

**Gapeng-北京大学-CV 2017-08-17 21:44:26**

这体现的是互补

**丁铭 清华 数据挖掘 2017-08-17 21:47:18**

我觉得看下来不说完美，但是至少是自恰的。就是凸集那地方没有理论保证😂

**丁铭 清华 数据挖掘 2017-08-17 21:49:02**

毕竟作者之一是我最崇拜的一个学长，捂脸（\*/∇＼\*）

**Gapeng-北京大学-CV 2017-08-17 21:49:43**

哈哈

**Gapeng-北京大学-CV 2017-08-17 21:49:48**

大家还有问题吗？

**Gapeng-北京大学-CV 2017-08-17 21:50:01**

没有的话，自由讨论吧

**Gapeng-北京大学-CV 2017-08-17 21:50:51**

这篇的讨论还可以继续，SCAN那篇我们以后再继续

**Guo-Jun Qi 2017-08-17 21:51:15**

P(x)是真实密度还是互补密度？

**Guo-Jun Qi 2017-08-17 21:51:43**

P\*是互补？

**Gapeng-北京大学-CV 2017-08-17 21:51:53**

p\*是互补

**Gapeng-北京大学-CV 2017-08-17 21:52:01**

p(x)是真实分布

**Guo-Jun Qi 2017-08-17 21:52:20**

那(4)里第二项不应是p\*吗？

**starif-西电-ml 2017-08-17 21:52:54**

求对数后 就差一个负号

**starif-西电-ml 2017-08-17 21:54:18**

如果用p\* 第二项应该也是- 而不是+吧？

**starif-西电-ml 2017-08-17 21:54:24**

还是我看错了

**Gapeng-北京大学-CV 2017-08-17 21:55:33**

（4）第二项是p

**泛艺术范儿-西电-DL 2017-08-17 21:55:56**

"(4)的前2项是KL(p\_G||p\*). 而p\*在(3)里面定义的，算出来就是那样吧，没啥问题。"

**Guo-Jun Qi 2017-08-17 21:56:23**

其实第二项就是在p上求g生成的样本的likelihood吧

**Guo-Jun Qi 2017-08-17 21:57:12**

Max这个likelihood不是max真实样本的概率？

**丁铭 清华 数据挖掘 2017-08-17 21:58:41**

min

**rwang-邓迪大学\_CV 2017-08-17 22:02:28**

还在读这篇文章。公式(5)里conditional entropy如何可以guarantee strong true-fake belief?

**泛艺术范儿-西电-DL 2017-08-17 22:02:50**

"我觉得作者的出发点是最小化KL(p\_G||p\*),而p\*可以看做p的“互补”。所以效果上看使得p\_G和p的“互补”接近。"

**Gapeng-北京大学-CV 2017-08-17 22:05:00**

true fake belief那个，就是毛豆大佬常说的，要求尖峰分布，任何一个都可以