|  |  |  |  |
| --- | --- | --- | --- |
| **文章信息** | **摘要及贡献** | **动机及方法概述** | **实验** |
| **阅读程度**：  方法精读  **文章标题：**  Contrastive Multiview Coding  **中文标题：**  对比多视角编码  **发表于：**  ECCV2020  **作者：**  Yonglong Tian  **单位：**  MIT CSAIL麻省理工计算机科学与人工智能实验室 | 人类通过许多感官通道来观察世界，例如，左眼看到长波光通道，或者右耳听到高频振动通道。每个视图都是含噪和不完整的，但重要的因素，如物理、几何和语义，往往在所有视图之间共享（例如，一只“狗”可以被看到、听到和感觉到）。（动机）  我们假设，一个强大的表征可以建模视图中的不变因素。基于这一假设，我们研究了一种对比编码方案，该方案学习了一种表示，旨在最大化不同观点之间的互信息，但在其他方面是紧凑的。（假说）（通过最大化视图之间的互信息来提取不同视图之间的共享信息）  我们的方法可以扩展到任意数量的视图（即提供可以处理了两个视图以上的扩展模型），并且与视图本身是什么没有关系。与基于预测学习或单视图重建的公式相比，所得到的学习表征在下游任务中具有最先进的性能，并随着更多视图的增加而改进。在Imagenet线性读出基准上，我们达到了68.4%的前1精度。（吹一吹效果）  1我们将对比学习应用于多视图设置，试图最大化同一场景的不同视图的表示之间的互信息（例如，不同图像通道或不同模态之间的互信息）。  2我们的方法产生的表现形式，在自我监督学习方面超过了最先进的模型。例如，在ImageNet线性读出评估中，我们达到了68.4%的前1位精度，略高于最先进的当前工作Bachman等人（2019）。  3我们表明，对比学习目标优于预测学习目标。  4我们扩展了框架，从两个以上的视图中学习，并表明学习的表示的质量随着视图数量的增加而提高。  5我们进行了控制实验来测量互信息对表示质量的影响。 | v z等字母的下标不同表示同一个样本的不同view，上标不同表示不是同一个样本。图片经过特征提取网络后获得其特征向量，注意每一中view对应一个特征提取网络。  定义一种距离度量：  通过一个权重向量去衡量两种view的特征向量之间的相似度。  定义损失函数：  ，以V1为锚点，遍历V2，例如以第一张图片的view1为锚点，第一张图片的view2即为正样本，其他图片的view2即为负样本。  ，当然也要对称过来以V2为锚点，遍历V1，由此形成两个view情况下的完整损失函数。  对于多个View的情况下分两种扩展形式：core view与full graph。      数字表示的是这部分信息被考虑的次数，例如b图中的数字6表示：V1V2V3V4四者共用的信息在计算L(V1,V2)，L(V1,V3)，L(V1,V4)，L(V2,V3)，L(V2,V4)，L(V3,V4)时被考虑到，总计被考虑了6次。  **损失函数Lconstrast中，分母难以计算，因此引入NCEloss：**    这个概率表示，已知与，与是同一个样本的不同view（D=1）的概率，其中分布在理想情况下，，是噪声分布，通常选择均匀分布即可，都是先验分布，公式是个后验分布。    最小化这个损失函数显然能够推进向理想的分布方向发展，又因为：，这时以神经网络的输出作为，就可以优化神经网络了。  本文也是采用memory bank的形式进行特征向量的存储，这与instance discrimination这篇论文一样。 | 实验部分没有细看，简单来说就是进行几个实验对前面贡献中提到的后三个点进行回应：  实验一将特征提取网络用于分类任务，简单来说是从一张图片中随机裁剪几张子图作为几个不同的view，以其中一个view为core view，使用Lc损失函数进行训练，最后使用core view对应的那个特征提取网络，将全连接层去掉后用于微调。验证了贡献二  实验二验证贡献四  实验三验证了贡献三  实验四验证了贡献五 |