Bootstrap Your Own Latent

A New Approach to Self-Supervised Learning

1. 摘要

我们介绍了Bootstrap Your Own Latent（BYOL），这是一种自我监督的图像表示学习的新方法。BYOL依赖于两个神经网络，被称为在线和目标网络，它们相互作用并相互学习。从一个图像的增强视图中，我们训练在线网络来预测同一图像在不同增强视图下的目标网络表示。同时，我们用在线网络的缓慢移动的平均值来更新目标网络。虽然最先进的方法依赖于负样本，但BYOL在没有负样本的情况下实现了SOTA。

1. 问题重述

许多不同的训练方法被提出来学习视觉表征，通常依赖于视觉借口任务。其中，最先进的对比方法[8, 9, 10, 11, 12]是通过减少同一图像的不同增强视图的表征之间的距离（"正对"），以及增加不同图像的增强视图的表征之间的距离（"负对"）来训练。这些方法需要仔细处理阴性对[13]，依靠大批量[8, 12]、内存库[9]或定制的挖掘策略[14, 15]来重新分辨阴性对。此外，它们的性能在很大程度上取决于对图像增强的选择[8, 12]。

BYOL不需要negative pairs，它对网络的输出进行迭代引导，以作为增强表示的目标。此外，与对比性方法相比，BYOL对图像增强的选择更为稳健；我们怀疑不依赖负数对是其稳健性提高的主要原因之一。虽然以前基于引导的方法使用了伪标签[16]、聚类指数[17]或少量的标签[18, 19, 20]，但我们建议直接引导表征。特别是，BYOL使用两个神经网络，被称为在线和目标网络，它们相互作用并相互学习。从一幅图像的增强视图开始，BYOL训练其在线网络来预测目标网络对同一图像的另一个增强视图的表示。虽然这个目标允许有折叠的解决方案，例如为所有图像输出相同的矢量，但我们的经验表明，BYOL并没有收敛到这样的解决方案。我们假设（见第3.2节），(i)在在线网络中增加一个预测器和(ii)使用在线参数的缓慢移动平均值作为目标网络的组合，鼓励在在线投影中编码越来越多的信息，避免崩溃的解决方案。

1. 详细阐述
2. 实验
3. 评析