# Model Optimization

Taylor Guo, 2020年12月08日 @ Biren上海

服务号二维码

目录

[GAN Understanding：理解GAN 1](#_Toc49784240)

[前言 1](#_Toc49784241)

[生成模型 2](#_Toc49784242)

[前言 2](#_Toc49784243)

[1 简介 2](#_Toc49784244)

[2 相关工作 5](#_Toc49784245)

[3  问题 6](#_Toc49784246)

[4  方法 6](#_Toc49784247)

[5  实现细节 8](#_Toc49784248)

[6  实验评估 8](#_Toc49784249)

[7  结论 8](#_Toc49784250)

[参考文献 9](#_Toc49784251)

[StyleGAN: 一种基于风格迁移的生成对抗网络的生成器 10](#_Toc49784252)

[前言 10](#_Toc49784253)

[生成模型 10](#_Toc49784254)

[DCGAN: Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks 11](#_Toc49784255)

[前言 11](#_Toc49784256)

[1 简介 11](#_Toc49784257)

[2 相关工作 12](#_Toc49784258)

## Auto-Precision Scaling for Distributed Deep Learning

<https://arxiv.org/abs/1911.08907>

<https://arxiv.org/pdf/1911.08907.pdf>

分布式训练的一个主要问题是梯度同步成为瓶颈，限制了训练的规模。低精度梯度可以减少带宽。

自动精度调整(Auto Precision Scaling, APS)算法可以在以低精度浮点值传递时改善准确度。它可以改善所有精度的准确度而使用较低的传递成本。图像分类和分割测试中用8位浮点梯度准确率降低<0.05%。8个Nvidia GTX1080Ti训练BERT/Wikipedia可能会占用40%的时钟时间。用低精度梯度会损害推断准确率，训练无法收敛。低精度容易下溢或上溢。有很多0之和INF值，训练无法收敛。

APS是一个高效的梯度传输基于网络层级的自适应算法。它可以是大批次训练在8位甚至4位符号类型收敛。

APS，网络层级自适应算法，改善低精度下的准确率；

可以在8节点的分布式系统中训练8bit的分类和分割模型；

可以在256节点的分布式系统中，8bit梯度训练ResNet-50；

可以应用于任意低精度浮点运算；

梯度同步优化

梯度同步有好几种方法:

特殊的网络拓扑结构：

多组三阶段实现all-reduce操作；

通常的做法是计算前一层网络的梯度，然后一起传输所有计算好的梯度。

梯度稀疏化

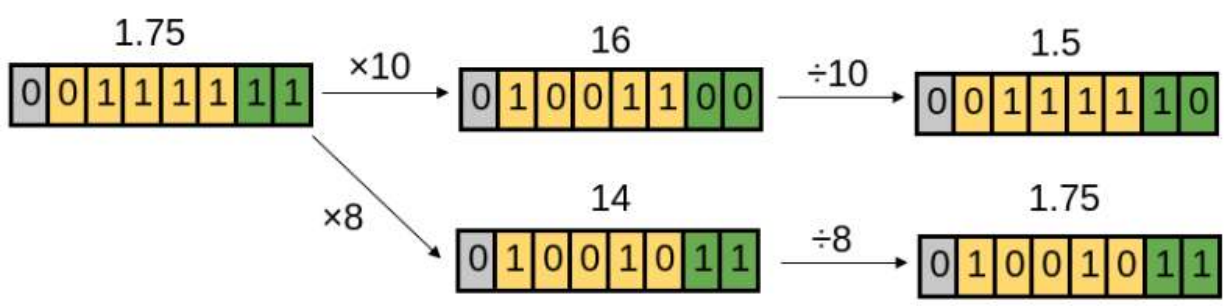
对梯度设置阈值，每次迭代只传输大于阈值的梯度；设置阈值的方法有测量每个梯度的重要性的L1范数；DGC方法累加局部梯度。

梯度量化：

缩减损失值来缩减梯度值

损失缩减

APS使用2的幂作为缩放因子，减少舍入误差。2的5次方



使用10和8作为缩放因子的区别，灰色部分是符号位，黄色部分是幂，绿色部分是底。因子是2的n次幂时，底是不变的，只是幂发生变化。

## DCGAN: Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks