

# 梯度优化裁剪(PCGrad)

Project Conflicting Gradients (PCGrad) [斯坦福, NIPS2020]

出自论文 Gradient Surgery for Multi-Task Learning, 这个名字十分形象: "Gradient Surgery".

在多任务训练期间, 如果能知道具体的梯度就可以利用梯度来动态更新  $\mathbf{w}$ 。如果两个任务的梯度存在**冲突**(即余弦相似度为负), 将任务A 的梯度投影到任务B 梯度的法线上。**即是消除任务梯度的冲突部分**, 减少任务间冲突。同时, **对模长进行了归一化操作**, 防止梯度被某个特别大的主导了。

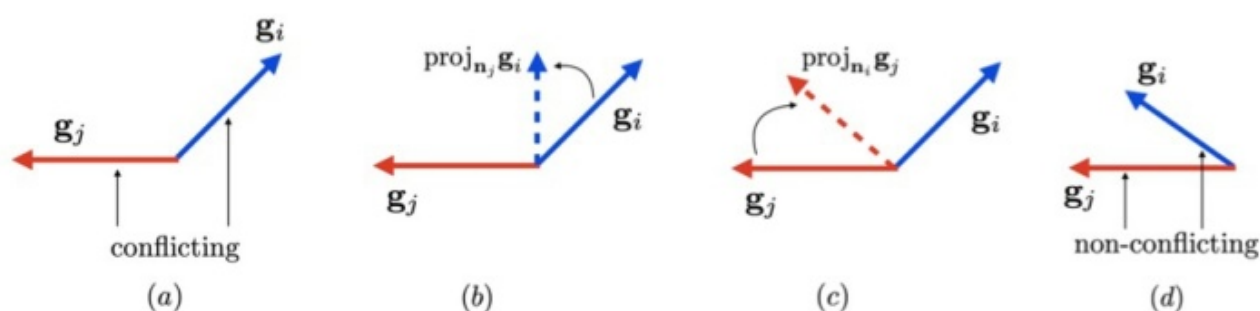


Figure 2: Conflicting gradients and PCGrad. In (a), tasks  $i$  and  $j$  have conflicting gradient directions, which can lead to destructive interference. In (b) and (c), we illustrate the PCGrad algorithm in the case where gradients are conflicting. PCGrad projects task  $i$ 's gradient onto the normal vector of task  $j$ 's gradient, and vice versa. Non-conflicting task gradients (d) are not altered under PCGrad, allowing for constructive interaction.

步骤:

1. 首先通过计算  $\mathbf{g}_i$  与  $\mathbf{g}_j$  之间的余弦相似度来判断  $\mathbf{g}_i$  是否与  $\mathbf{g}_j$  冲突; 其中负值表示梯度冲突。
2. 如果余弦相似度是负数, 我们用它在  $\mathbf{g}$  的法线平面上的投影替换。如果梯度不冲突, 即余弦相似度为非负, 原始梯度  $\mathbf{g}_i$  保持不变。

$$\mathbf{g}_j : \mathbf{g}_i = \mathbf{g}_i - \frac{\mathbf{g}_i \mathbf{g}_j}{\|\mathbf{g}_j\|^2}$$

其实, 这篇文章是很有问题的。梯度冲突不一定是个坏事, 而可以带来正则化的好处, 如果只是把冲突梯度完全抹去恐有不妥。而且, 文章并没有对梯度裁剪进行消融实验。万一是梯度裁剪、而不是Project conflict 导致的性能提升呢? 不过, 这个方法在我的测试上也是表现比原来的好的。

