

使用不确定性来衡量场景集合和语义的损失的多任务学习

Alex Kendall, Yarin Gal, Roberto Cipolla

Apr 2018

摘要

许多深度学习应用从拥有多个回归和分类目标的多任务学习中获益。在这篇文章中我们发现这种系统的性能严重依赖于每个任务损失间的相对权重。手动调整这些权重是困难并且昂贵的，这使得多任务学习在实际中受到限制。我们通过考虑每个任务的同方差不确定性，为多任务深度学习提出了一种基于一系列规则的方式来衡量多种损失函数。这使得我们可以同时在分类和回归的设定下，学习不同单位或者大小的数值。我们通过在单目输入图片中学习像素点深度回归、语义和实例分割来阐明我们的模型。或许令人惊讶的是，我们展示我们的模型可以学习多个任务的权重，同时性能超过为每个任务分别训练单独的模型。

1 Introduction

多任务学习旨在通过从共享的表示来学习多个目标来提升学习效率和预测精度。多任务学习在多个机器学习应用中都十分流行，从计算机视觉到自然语言处理再到语音识别。

我们在计算机视觉中的视觉场景理解设定下探索多任务学习。场景理解算法必须同时理解场景的几何和语义。这形成了一个有趣的多任务学习问题，因为场景理解包含同时学习有不同单位和尺度的多个回归和分类任务。视觉场景理解的多任务学习在禁止长时间计算的系统，例如机器人中使用的系统，中是至关重要的。将所有任务结合为单个模型减少了计算并允许那些系统实时运行。

之前的同时学习多个任务的方法使用简单的损失加权和，其中权重是均匀的或者是手动调整的。然而，我们表明性能高度决定于合适地选择每个任务

损失间的权重。寻找最优权重昂贵到不可接受，同时手动调整解决是困难的。我们发现每个任务的最优权重取决与测量的尺度（例如单位，分米还是微米）以及最终，任务噪声的数量级。

在这个工作中，我们利用同方差不确定性提出了一种基于一系列规则的方式来结合多个损失函数来同时学习多个目标。我们将同方差不确定性阐述为任务相关的权重，同时展示如何导出一个基于规则的可以平衡不同回归和分类损失的多任务损失函数。与为每个任务分别学习相比，我们的方法可以学习最优地平衡这些权重，进而得到更好的性能。

特别的，我们通过三个学习场景几何和语义的任务来阐述我们的方法。首先，我们学习在像素级别进行物体分类，这也被称为语义分割。第二，我们的模型会进行实例分割，这是一个更难的任务，它为图片中的每一个物体分割单独的 mask（例如，一个单独的，精细的 mask 为每一个路上的车）。由于它不仅要求估计每一个像素的类别，还要估计这个像素属于哪个物体，所以这是一个比语义分割更难的任务。它同样也比通常只预测物体边界框的检测任务更加复杂。最后，我们的模型预测像素级别的单位深度。