Focal-Loss

深度学习几分钟精通系列

范仁义

一、课件资料

- 原论文地址：

  [1708.02002] Focal Loss for Dense Object Detection
  https://arxiv.org/abs/1708.02002

- 论文、论文翻译、代码、课件及各种资料位置：

  fry404006308/fry_course_materials: 范仁义录播课资料,
  https://github.com/fry404006308/fry_course_materials

二、为何有focal-loss

解决样本极度不均衡的问题

3. Focal Loss

The *Focal Loss* is designed to address the one-stage object detection scenario in which there is an extreme imbalance between foreground and background classes during training (*e.g.*, 1:1000). We introduce the focal loss starting from the cross entropy (CE) loss for binary classification[1]:

Abstract

The highest accuracy object detectors to date are based on a two-stage approach popularized by R-CNN, where a classifier is applied to a sparse *set of candidate object locations. In contrast, one-stage detectors that are applied over a regular,* dense *sampling of possible object locations have the potential to be faster and simpler, but have trailed the accuracy of two-stage detectors thus far. In this paper, we investigate why this is the case. We discover that the extreme foreground-background class imbalance encountered during training of dense detectors is the central cause. We propose to address this class imbalance by reshaping the standard cross entropy loss such that it down-weights the loss assigned to well-classified examples. Our novel* Focal Loss *focuses training on a sparse set of hard examples and prevents the vast number of easy negatives from overwhelming the detector during training. To evaluate the effectiveness of our loss, we design and train a simple dense detector we call RetinaNet. Our results show that when trained with the focal loss, RetinaNet is able to match the speed of previous one-stage detectors while surpassing the accuracy of all existing state-of-the-art two-stage detectors. Code is at:* https://github.com/facebookresearch/Detectron.

三、样本不均衡带来的问题

Class Imbalance: Both classic one-stage object detection methods, like boosted detectors [37, 5] and DPMs [8], and more recent methods, like SSD [22], face a large class imbalance during training. These detectors evaluate 10^4-10^5 candidate locations per image but only a few locations contain objects. This imbalance causes two problems: (1) training is inefficient as most locations are easy negatives that contribute no useful learning signal; (2) en masse, the easy negatives can overwhelm training and lead to degenerate models. A common solution is to perform some form of hard negative mining [33, 37, 8, 31, 22] that samples hard examples during training or more complex sampling/reweighing schemes [2]. In contrast, we show that our proposed focal loss naturally handles the class imbalance faced by a one-stage detector and allows us to efficiently train on all examples without sampling and without easy negatives overwhelming the loss and computed gradients.

四、focal-loss函数长啥样

1、交叉熵损失函数

$$\mathrm{CE}(p, y) = \mathrm{CE}(p_t) = -\log(p_t).$$

2、focal-loss

$$\mathrm{FL}(p_t) = -\underbrace{(1 - p_t)}_{\text{权重}}{}^{\gamma} \log(p_t).$$

五、focal-loss核心思想

3.2. Focal Loss Definition

As our experiments will show, the large class imbalance encountered during training of dense detectors overwhelms the cross entropy loss. Easily classified negatives comprise the majority of the loss and dominate the gradient. While α balances the importance of positive/negative examples, it does not differentiate between easy/hard examples. Instead, we propose to reshape the loss function to down-weight easy examples and thus focus training on hard negatives.

六、focal-loss为什么有用

1、交叉熵损失函数

$$\mathrm{CE}(p, y) = \mathrm{CE}(p_t) = -\log(p_t).$$

A、当我某次预测为0.2时，显然预测的非常不准，是困难样本，这时对应的损失为：
CELoss = -ln(0.2) = 1.609
B、当我某次预测为0.9时，显然预测的比较准确，是简单样本，这时对应的损失为：
CELoss = -ln(0.9) = 0.105

2、focal-loss

$$\mathrm{FL}(p_t) = -(1 - p_t)^{\gamma} \log(p_t).$$

假设γ=2
A、当我某次预测为0.2时，显然预测的非常不准，是困难样本，这时对应的损失为：
FLoss = -(1-0.2)^2*ln(0.2) = 0.64*1.609
B、当我某次预测为0.9时，显然预测的比较准确，是简单样本，这时对应的损失为：
FLoss = -(1-0.9)^2*ln(0.9) = 0.01*0.105

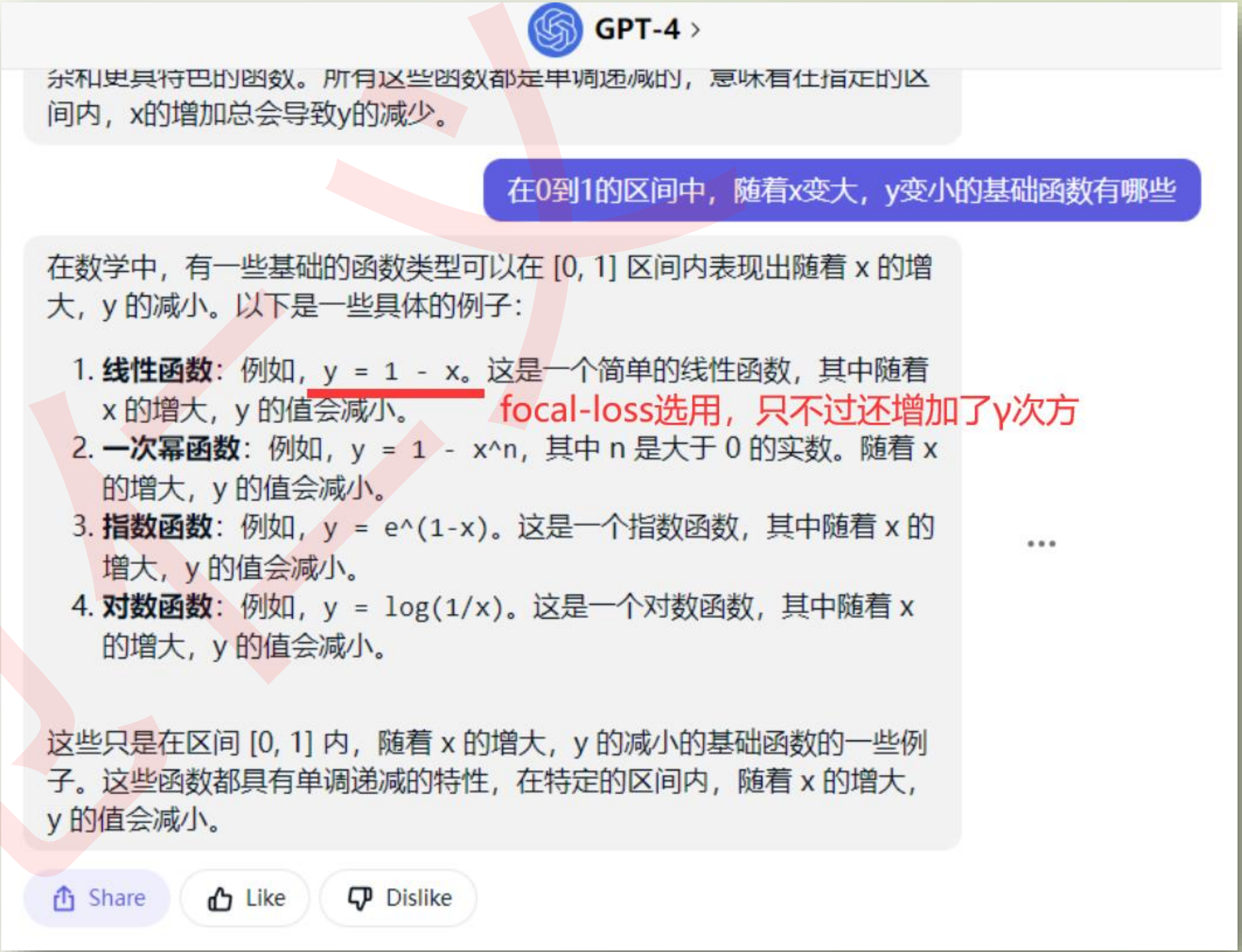
结论：应用focal-loss后，简单样本的损失缩小了100倍，困难样本的损失缩小了2倍不到，也就是降低了简单样本权重，增加了困难样本权重

七、知道原理，轻松设计损失函数

- Focal-loss原理：降低简单样本权重，增加困难样本的权重
- 简单样本：预测值Pt较高的，比如0.9
- 困难样本：预测值Pt较低的，比如0.2
- 设计：损失 = 权重函数 * (-log(pt))
- 自己设计原理：找一个权重函数，让自变量Pt在[0,1]的范围内，Pt越大，对应的值越小

focal-loss

$$\mathrm{FL}(p_t) = -(1-p_t)^{\gamma}\log(p_t).$$



八、focal-loss代码详解

- 原focal-loss

$$\mathrm{FL}(p_t) = -(1-p_t)^{\gamma}\log(p_t).$$

- focal-loss增加a

a的作用是给不同的分类不同的权重值

In practice we use an α-balanced variant of the focal loss:

$$\mathrm{FL}(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t). \tag{5}$$

We adopt this form in our experiments as it yields slightly improved accuracy over the non-α-balanced form. Finally, we note that the implementation of the loss layer combines the sigmoid operation for computing p with the loss computation, resulting in greater numerical stability.