

F - measure 总结

Precision (查准率) :

$$Precision = \frac{TP}{TP + FP}$$

Recall (查全率) :

$$Recall = \frac{TP}{TP + FN}$$

TP: True Positives

FP: False Positives

TN: True Negatives

FN: False Negatives

我们都希望 Precision 和 Recall 越高越好，最好能够达到 $P = R = 1$ ，然而这只是一种理想情况。如果我们希望：被检索到的内容越多越好，这是追求“查全率”；如果我们希望：检索到的文档中，真正想要的、也就是相关的越多越好，不相关的越少越好，这是追求“准确率”。

问题：

Precision 和 Recall 的组合能够反应分割效果的好坏：比如通过计算得到一个模型的 $P = 1$ ， $R = 0.01$ ，那么就可以说明该模型的分割性能不佳。但 Precision 和 Recall 在对比时往往会出现问题：假设有两个模型， $P = 0.9$ ， $R = 0.19$ ； $P = 1$ ， $R = 0.09$ ，那么两个模型相比，哪个模型更好？因此我们需要一种方式来综合考量 Precision 和 Recall，所以才有了 F-measure 的产生。

F-measure :

$$F_{\beta} = (1 + \beta^2) \frac{Precision \cdot Recall}{\beta^2 \cdot Precision + Recall}$$

通过调整 β ，可以改变Precision和Recall在F-measure中的权重。当 $\beta < 1$ 时，Precision权重大于Recall；当 $\beta > 1$ ，Recall权重大于Precision。在实际情况中，根据需求来确定 β 的值。如果是做搜索，那就是保证查全的情况下提升准确率；如果做疾病监测、反垃圾，则是保准确率的条件下，提升查全。具体来说，这里举个例子：如果是做显著性检测，Precision所占比重应该大于Recall，如果是做视觉追踪，则应该相反。

当 $\beta = 1$ 时，Precision 和 Recall 在 F-measure 中所占比重相等，这就是平常所说的 F1-measure。

F1-measure：

$$F_1 = 2 \frac{Precision \cdot Recall}{Precision + Recall}$$

1. PR 曲线的产生？

因为 Precision、Recall 是针对二类分类问题的。对于普通的二类分割（只有一个目标），每张图像只能产生一个 P 和一个 R 值，因此无法画出 PR 曲线。而对于含有参数的分类问题，可根据调整参数（如阈值等）来确定约束。不同的参数对应 PR 曲线上的一个点，依次选取不同的参数可以得到不同的点从而画出完整的曲线。对于整个数据集，只需分别在不同的参数下对整个数据集求取均值来获取不同的点，最终可以得到整个数据集的 PR 曲线。

2. 单从 F-measure 值是否可以直接评判分割性能的好坏？

采用不同的分割方式，P、R 可以是 $[0, 1]$ 上的任意值。比如可能会得到两个相等的 F-measure（ $P = 1/2, R = 1/4$ ； $P = 1, R = 1/5$ ），在这种情况下还要再考虑二者分别对应的 P、R 值。