

召回率与精确率之间的区别

召回率与精确率之间的区别

概述

定义

举个栗子

参考文献

概述

刚刚接触文本挖掘领域，经常会碰到召回率（recall）和精确率（precision）这两个专业术语，总是不能很清楚的记得，本文将给出这两个概念的定义。

定义

精确率是针对**预测结果**而言的，它表示的是预测为正的样本中有多少是真正的样本。那么预测为正就有两种可能，一种是正类样本被预测为正类（TP），另一种是将负类样本预测是正类（FP），即

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

而召回率是针对我们原来的样本而言，它表示的是样本中的正类有多少被预测正确了，该种情况下也有两种可能，一种是原来的正类被预测为正类（TP），另一种是原来的正类被预测为负类（FN），即

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

其实看上面的两个公式，仅仅是公式的分母不同，一个分母是预测为正的样本数，另一个是原来样本中所有的正样本数。

举个栗子

假设我们手上有60个正样本，40个负样本，我们要找出所有的正样本，系统查找出50个，其中有40（TP）个是真正的正样本，那么就存在10（FP）个是负样本被错误预测为正样本。

按照上面的公式能够得出：

- 精确率(precision) = $TP/(TP+FP) = 80\%$
- 召回率(recall) = $TP/(TP+FN) = 2/3$

参考文献

1. <https://www.zhihu.com/question/19645541>
2. 机器学习性能评估指标 <https://charleshm.github.io/2016/03/Model-Performance/>