

Mean Average Precision- mAP 評估測量方法



Sam Zheng

Sep 25, 2018 · 4 min read

最近在看有關object detection的論文, 會看到一個評估測量的方法mAP, mean average precision,

網路上有很多例子, 但可能我資質駑鈍我真的很多都看不懂, 在看了此篇之後終於了解了一些, 我想舉一個更適合表達的例子來描述這個評估測量

今天假設有十台車, 裡面有4台, 是你喜歡並且想要的車, 我們姑且就把它用+1標記, 剩下都是你沒有那麼愛的, 用-1標記, 所以標記到的結果是

+1, +1, +1, +1, -1, -1, -1, -1, -1, -1

接著, 將這十台車的資訊丟給機器幫你決定, 它會列出排名, 說明你最喜歡a車第一名, b車第二名.....以此類推, 排出十台車的順位, 假設排出的順位是(最前面代表機器認為你最喜歡那台車)

+1, -1, +1, +1, -1, -1, -1, -1, -1, +1

接著我們要來看這台機器是否能準確的預測我們對於車子的喜好, 所以我們開始審查機器所排列的車, 我們先看第一台

是 +1

很好, 看來機器覺得我最喜歡的車剛好真的我也蠻喜歡的, 那機器的第二順位是?

-1

恩.... 但這台車我不太愛, 機器真的懂我嗎? 來看第三台

+1 不錯, 這台也是我很喜歡的車, 那第四台呢?

+1 挨優, 機器看起來還不錯耶!!

由上面可以得知, 你對於這機器的準確率, 會隨著你往後面的順位看越多, 而有所改觀, 所以這邊我們要進入一個觀念, **recall** 與 **precision** 這觀念

假設你看到第四台就不在往下看了

recall 代表你喜歡的车, 並且被機器選中除以全部你喜歡的车 = $3/4$

precision 代表 你喜歡的车, 並且被機器選中除以你只看到機器幫你分的前幾名, 剛剛你只看到第四個就不看了, 所以這邊會等於 $2/4$

這邊還有一個概念, 通常**recall**越大 **precision**會越小, 我們可以注意到當**recall**越高代表很可能你全部喜歡的车很少, 機器很難幫你挑, 這邊在舉一個極端的例子, 假使有一千台車你只愛2台車, 假設機器幫你找到了並將排名一台放在最前面, 另一台放在後面, 而這時我只在乎它的第一台車有沒有準, **recall**就是 $1/2$, 而**precision**是 $1/1$, 可是當你看到第二台, 第三台時, **recall**都沒有變, 但你會發現**precision**卻一直的下降 例如我看到第20台車, **precision**就是 $1/20$

這樣對於**recall** 與 **precision**有沒有多一些些的理解呢

最後我們要看到要如何去評估這台機器的好壞, 這邊們就要回到最前面的例子囉, 只有時台車的

我們做個表格幫助我們判斷

機器排名	我的喜歡	recall值	presicion值
1	+1	$1/4$	$1/1$
2	-1	$1/4$	$1/2$
3	+1	$2/4$	$2/3$
4	+1	$3/4$	$3/4$
5	-1	$3/4$	$3/5$
6	-1	$3/4$	$3/6$
7	-1	$3/4$	$3/7$
8	-1	$3/4$	$3/8$
9	-1	$3/4$	$3/9$
10	+1	$4/4$	$3/10$

我們可以發現總共只有四組的recall, 分別是 1/4, 2/4, 3/4, 4/4

我們分別去找到這些recall對應到最大的precision

1/4對應到最大的precision為1/1, 看上方的第1列

2/4對應到最大的precision為2/3, 看上方的第3列

3/4對應到最大的precision為3/4, 看上方的第4列

4/4對應到最大的precision為3/10, 看上方的第10列

最後 將這些全部相加 除以你愛車子的總量, 即可得到

$$(1 + 2/3 + 3/4 + 3/10) / 4 = 0.6775$$

這就是你的Average precision(AP)

那多了個 mean呢? 別忘了我們只有判斷你對你的車子喜不喜歡而已, 如果這時後來有顏色, 廠牌等等的因素有著不同的AP

那在將這些ap加起來除以所有的因素數量

就是要計算的mAP了

如有觀念上的錯誤, 還請大家多多提醒了

[About](#) [Help](#) [Legal](#)

Get the Medium app

