I am Learning

121,875

Sunday, 21 August 2011

Micro- and Macro-average of Precision, Recall and F-Score

I posted several articles explaining how precision and recall can be calculated, where F-Score is the equally weighted harmonic mean of them. I was wondering- how to calculate the average precision, recall and harmonic mean of them of a system if the system is applied to several sets of data.

Tricky, but I found this very interesting. There are two methods by which you can get such average statistic of information retrieval and classification.

1. Micro-average Method

In Micro-average method, you sum up the individual true positives, false positives, and false negatives of the system for different sets and the apply them to get the statistics. For example, for a set of data, the system's

True positive (TP1)= 12
False positive (FP1)=9
False negative (FN1)=3

Then precision (P1) and recall (R1) will be 57.14 and 80

and for a different set of data, the system's

True positive (TP2)= 50
False positive (FP2)=23
False negative (FN2)=9

Then precision (P2) and recall (R2) will be 68.49 and 84.75

Now, the average precision and recall of the system using the Micro-average method is

Micro-average of precision = (TP1+TP2)/(TP1+TP2+FP1+FP2) = (12+50)/(12+50+9+23) = 65.96
Micro-average of recall = (TP1+TP2)/(TP1+TP2+FN1+FN2) = (12+50)/(12+50+3+9) = 83.78

The Micro-average F-Score will be simply the harmonic mean of these two figures.

2. Macro-average Method

The method is straight forward. Just take the average of the precision and recall of the system on different sets. For example, the macro-average precision and recall of the system for the given example is
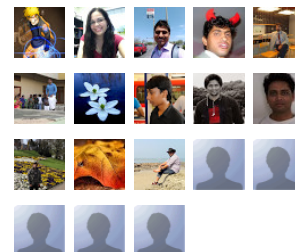
Macro-average precision = (P1+P2)/2 = (57.14+68.49)/2 = 62.82
Macro-average recall = (R1+R2)/2 = (80+84.75)/2 = 82.25

The Macro-average F-Score will be simply the harmonic mean of these two figures.

Followers

Followers (18)

[Follower avatars]

Follow

Blog Archive

About Me

Rushdi Shams

View my complete profile

Suitability

Macro-average method can be used when you want to know how the system performs overall across the sets of data. You should not come up with any specific decision with this average.

On the other hand, micro-average can be a useful measure when your dataset varies in size.

Posted by Rushdi Shams at 14:08　　　　　　　 G+

Labels: macro-average f-score, macro-average precision, macro-average recall, micro-average f-score, micro-average precision, micro-average recall, Precision, Recall

10 comments:

Mohamed Yahya 23 January 2012 at 01:11

Thanks for the article, just a minor typo: second to last paragraph: Micro rather than macro.

Reply

　　Replies

　　Rushdi Shams　23 January 2012 at 04:41

　　Thanks a lot, Yahya. I corrected it. But the typo is in the last paragraph. If your dataset varies in size, micro-average is the useful tool.

　　sue 9 December 2012 at 12:01

　　hi, thank you for the article, please i have a general question related to learning algorithms. I found this expression in an article : The training data for the SVM classi ers are highly unbalanced as the proportion of positive training web pages ranges from 2% to 18%.
　　What do we mean by unbalanced data? and by positive/negative training?
　　ReplyDelete

　　Reply

LylyKim 23 May 2012 at 08:59

thanks!

Reply

sue 9 December 2012 at 11:59

This comment has been removed by the author.

Reply

Pallika 6 March 2015 at 06:26

I think there's a problem with your macro-average precision calculation. the denominator should be the sum of true positives and false positives. The micro average for precision and recall are the same. See here:
http://metaoptimize.com/qa/questions/8284/does-precision-equal-to-recall-for-micro-averaging

Reply

Rushdi Shams 6 March 2015 at 08:28

@Pallika. The scenarios are different. In my case, one classifier is applied on two different datasets. The number of instances are different in two datasets. But the example you provided is classifier performance for three different class labels. So the sum of positive and negative is always 27. The two macro and micro evaluation is different.

Reply

krishnan 19 June 2015 at 04:07

Hi..Thank you for the article. I did some exparimentation on Text classification which consists 7 (labled documents prepared on my own)ctegories of documents.I did too much stemming and i got average precision 57%, micro precision=macro precision=1,micro recall=macro recall=0.please suggests me something to improve results(like bad training set,light stemming, threshold value and threshold step size modification, k if classifier is knn,how to improve recall for better F1 score).
Thanks.

Reply

vignesh G 21 April 2017 at 18:28

Good article.... if you explain what is micro and macro average in 3 lines it will be more hepful for readers...

Reply

Abdol 12 November 2017 at 15:35

This comment has been removed by the author.

Reply

```
Enter your comment...
```

Comment as: Select profile... ▼

Publish   Preview

Newer Post   Home   Older Post

Subscribe to: Post Comments (Atom)

Simple theme. Powered by Blogger.