Calculating Precision, Recall, and F1-score of a Multi-class Classifier, like- a Part-of-Speech (PoS) Tagger

CS 565: Intelligent Systems and Interfaces
Indian Institute of Technology Guwahati
Jan-May 2016

Precision, recall, and F1-score are used to measure the performance of a classifier. The higher these values, the better the classifier.

Assume, the input dataset contains a total of 492 word tokens: (176 articles + 201 adjectives + 115 nouns). Hence, the given set of PoS tags is $\{AT, JJ, NN\}$. 'AT', 'JJ', 'NN' stand for Article, Adjective, and Singular noun, respectively. Once, our classifier completes execution, the following steps need to be followed.

1 Step 1: Confusion Matrix

Let, Confusion Matrix $C = \begin{array}{c} AT_{True} \\ JJ_{True} \\ NN_{True} \end{array} \begin{pmatrix} AT_{Classified} & JJ_{Classified} & NN_{Classified} \\ 50 & 37 & 89 \\ 90 & 45 & 66 \\ 11 & 9 & 95 \end{pmatrix}$

The diagonal elements of C imply number of correct classifications.

For example, $C_{AT_{True},\ AT_{Classified}} = 50 \Rightarrow$ Number of word tokens that are classified as 'Article' and truly are articles is 50.

On the other hand, the off-diagonal elements of C imply number of incorrect classifications. For example, $C_{AT_{True},\ JJ_{Classified}} = 37 \Rightarrow$ Number of word tokens that are classified as 'Adjective' but are articles in truth is 37.

2 Step 2: True Positive (TP), False Positive (FP), False Negative (FN) for Each Individual Class

From the confusion matrix C, we now calculate TP, FP, and FN for every class/PoS tag. Let us start with the PoS tag 'AT'.

TP_{AT} = Number of word tokens that our classifier correctly classified as 'Article'.

FP_{AT} = Number of word tokens that our classifier incorrectly classified as 'Article'.

FN_{AT} = Number of word tokens that are articles in truth but our classifier incorrectly classified them with some other PoS tag.

$$
\begin{aligned}
TP_{AT} &= C_{AT_{True},\ AT_{Classified}} \\
&= 50
\end{aligned}
\tag{1}
$$

$$FP_{AT} = C_{JJ_{True},\ AT_{Classified}} + C_{NN_{True},\ AT_{Classified}}$$
$$= 90 + 11$$
$$= 101 \qquad\qquad (2)$$

$$FN_{AT} = C_{AT_{True},\ JJ_{Classified}} + C_{AT_{True},\ NN_{Classified}}$$
$$= 37 + 89$$
$$= 126 \qquad\qquad (3)$$

Similarly, we can find $TP_{JJ}, \ldots, FN_{JJ}, TP_{NN}, \ldots, FN_{NN}$.

3 Step 3: Precision, Recall, F1-score for Each Individual Class

In this Section, we discuss how to compute precision, recall, F1-score of our classifier for a given class. In Section 4, we shall discuss how to compute overall precision, recall, F1-score of the classifier, which is our ultimate objective. Again, let us start with the PoS tag 'AT'.

PRE_{AT} = Precision of our classifier for the PoS tag 'AT'.
REC_{AT} = Recall of our classifier for the PoS tag 'AT'.
$F1_{AT}$ = F1-score of our classifier for the PoS tag 'AT'.

$$PRE_{AT} = \frac{TP_{AT}}{TP_{AT} + FP_{AT}}$$
$$= \frac{50}{50 + 101} \qquad From\ Eqn\ (1)\ and\ (2)$$
$$= \frac{50}{151}$$
$$= 0.3311 \qquad\qquad (4)$$

$$REC_{AT} = \frac{TP_{AT}}{TP_{AT} + FN_{AT}}$$
$$= \frac{50}{50 + 126} \qquad From\ Eqn\ (1)\ and\ (3)$$
$$= \frac{50}{176}$$
$$= 0.2841 \qquad\qquad (5)$$

$$F1_{AT} = 2 \times \frac{PRE_{AT} \times REC_{AT}}{PRE_{AT}\ +\ REC_{AT}}$$
$$= 2 \times \frac{0.3311 \times 0.2841}{0.3311\ +\ 0.2841} \qquad From\ Eqn\ (4)\ and\ (5)$$
$$= 2 \times \frac{0.0941}{0.6152}$$
$$= 2 \times 0.153$$
$$= 0.306 \qquad\qquad (6)$$

4 Step 4: Precision, Recall, F1-score of a Classifier

4.1 Precision of a Classifier

Here, PRE stands for Precision. It can be of two types. They are: Micro average (PRE_{micro}) and Macro average (PRE_{macro}).

$$PRE_{micro} = \frac{TP_{AT} + TP_{JJ} + TP_{NN}}{TP_{AT} + TP_{JJ} + TP_{NN} + FP_{AT} + FP_{JJ} + FP_{NN}} \tag{7}$$

$$
\begin{aligned}
PRE_{macro} &= \frac{PRE_{AT} + PRE_{JJ} + PRE_{NN}}{Number\ of\ Classes} \\
&= \frac{PRE_{AT} + PRE_{JJ} + PRE_{NN}}{|\{AT, JJ, NN\}|} \\
&= \frac{PRE_{AT} + PRE_{JJ} + PRE_{NN}}{3}
\end{aligned} \tag{8}
$$

4.2 Recall of a Classifier

Here, REC stands for Recall. It can be of two types. They are: Micro average (REC_{micro}) and Macro average (REC_{macro}).

$$REC_{micro} = \frac{TP_{AT} + TP_{JJ} + TP_{NN}}{TP_{AT} + TP_{JJ} + TP_{NN} + FN_{AT} + FN_{JJ} + FN_{NN}} \tag{9}$$

$$
\begin{aligned}
REC_{macro} &= \frac{REC_{AT} + REC_{JJ} + REC_{NN}}{Number\ of\ Classes} \\
&= \frac{REC_{AT} + REC_{JJ} + REC_{NN}}{|\{AT, JJ, NN\}|} \\
&= \frac{REC_{AT} + REC_{JJ} + REC_{NN}}{3}
\end{aligned} \tag{10}
$$

4.3 F1-score of a Classifier

Here, $F1$ stands for F1-score. It is again of two types: Micro average ($F1_{micro}$) and Macro average ($F1_{macro}$).

$$F1_{micro} = 2 \times \frac{PRE_{micro} \times REC_{micro}}{PRE_{micro} + REC_{micro}} \tag{11}$$

$$F1_{macro} = 2 \times \frac{PRE_{macro} \times REC_{macro}}{PRE_{macro} + REC_{macro}} \tag{12}$$

5 References

- http://sebastianraschka.com/faq/docs/multiclass-metric.html

- https://www.youtube.com/watch?v=OwwdYHWRB5E