Why Precision, Recall, and Accuracy – and Not Some Other Characteristics?

David A. Davis, Nestor A. Escobedo, Laura A. Guevara,
Brian G. Rodiles Delgado, Alejandro Rodriguez, and Vladik Kreinovich
Department of Computer Science, University of Texas at El Paso
500 W. University, El Paso, TX 79968, USA
dadavis3@miners.utep.edu, naescobedo@miners.utep.edu, laguevara3@miners.utep.edu,
bgrodilesde@miners.utep.edu, arodriguez326@miners.utep.edu, vladik@utep.edu

1. Formulation of the problem

- Classification methods are often not perfect.
- In addition to true positive (TP) and true negative (TN) cases, we also have false positive (FP) and false negative (FN) cases.
- To gauge the quality of a classification method, we need to take into account the numbers of all these four categories.
- In principle, we can have many different combinations of these four numbers.
- Empirically, the following three combinations are most frequently used:
  - precision $P = TP/(TP + FP)$,
  - recall $R = TP/(TP + FN)$, and
  - accuracy $A = (TP + TN)/(TP + TN + FP + FN)$.
- We also use their combination $F1 = P \cdot R/(P + R)$.
- A natural question is: why these characteristics and not others?

2. Towards an explanation

- Each correct classification brings benefits, each false classification brings losses.
- The method should be applied if the benefits are larger than the losses.
- With respect to benefits:
  - sometimes, benefits b_{TP} and b_{TN} of TP and TN are similar, and
  - sometimes, one of them brings more benefits.
- For example:
  - detecting cancer may save a life, while
  - correctly identifying a non-cancerous tumor simply saves a patient from a few further procedures.
- In principle, we could have three cases: $b_{TP} \approx b_{TN}$, $b_{TP} \gg b_{TN}$, and $b_{TN} \gg b_{TP}$.

3. Towards an explanation (cont-d)

- If TN brings more benefits, we can simply rename negative to positive.
- So we have two cases: $b_{TP} \approx b_{TN}$ and $b_{TP} \gg b_{TN}$.
- In the first approximation, when we ignore small numbers and small differences:
  - the first case means $b_{TP} = b_{TN}$, and
  - the second means $b_{TP} > 0$ and $b_{TN} = 0$.
- Similarly for losses, in the first approximation, we can consider three possible cases:
  - the case when $\ell_{FP} = \ell_{FN}$,
  - the case when $\ell_{FP} > 0$ and $\ell_{FN} = 0$, and
  - the case when $\ell_{FN} > 0$ and $\ell_{FP} = 0$.

4. Towards an explanation (cont-d)

- Let us first consider the case when $b_{TP} > 0$, $b_{TN} = 0$, $\ell_{FP} > 0$ and $\ell_{FN} = 0$.
- In this case, the method is beneficial if $b_{TP} \cdot TP > \ell_{FP} \cdot FP$, i.e., equivalently, when

$$r_1 \stackrel{\text{def}}{=} TP/FP > \ell_{FP}/b_{TP}.$$

- The larger the ratio r_1, the more cases when this method is useful.
- So, the quality of the method is larger if the ratio r_1 is larger.
- Alternatively, we can take any strictly increasing function of r_1, e.g., $1/(1+1/r_1)$.
- Applying this function to $r_1 = TP/FP$, we get exactly the precision – which explains why precision is used.

5. Towards an explanation (cont-d)

- In the case when $b_{TP} > 0$, $b_{TN} = 0$, $\ell_{FN} > 0$, and $\ell_{FP} = 0$, a similar argument leads to recall.
- In the case when $b_{TP} = b_{TN}$ and $\ell_{FP} = \ell_{FN}$, a similar argument leads to accuracy.

6. Why only three?

- For general values of benefits and losses, the method is effective if

$$b_{TP} \cdot TP + b_{TN} \cdot TN > \ell_{FP} \cdot FP + \ell_{FN} \cdot FN.$$

- If we divide both sides by TP, we get an equivalent inequality

$$b_{TP} + b_{TN} \cdot TN/TP > \ell_{FP} \cdot FP/TP + \ell_{FN} \cdot FN/TP.$$

- This inequality has three unknown ratios

$$R_1 \stackrel{\text{def}}{=} FP/TP, \quad R_2 \stackrel{\text{def}}{=} FN/TP, \text{ and } R_3 \stackrel{\text{def}}{=} TN/TP.$$

- One can check that:
  - by dividing both the numerator and the denominator of the expressions for P, R, and A by TP,
  - that these three values depend only on these three ratios:

$$P = 1/(1+R_1), \; R = 1/(1+R_2), \text{ and } A = (1+R_3)/(1+R_1+R_2+R_3).$$

7. Why only three (cont-d)

- Thus, when we know the values of P, R, and A, we have 3 equations from which we can determine all three unknown ratios:

$$R_1 = 1/P - 1, \quad R_2 = 1/A - 1, \text{ and}$$

$$R_3 = (A \cdot (R_1 + R_2 + 1) - 1)/(1 - A).$$

- Hence, once we know P, R, and A, we will be able to predict:
  - for each combination of benefits and losses,
  - whether this method is applicable.
- So, the three characteristics are indeed sufficient – all other characteristics can be described in terms of these three (just like F1).

8. Acknowledgments

This work was supported in part:

- by the US National Science Foundation grants:
  - 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science),
  - HRD-1834620 and HRD-2034030 (CAHSI Includes),
  - EAR-2225395 (Center for Collective Impact in Earthquake Science C-CIES),
- by the AT&T Fellowship in Information Technology, and
- by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Focus Program SPP 100+ 2388, Grant Nr. 501624329,