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

Abstract In principle, there can be many different characteristics of classification quality. However, in practice, mostly the following three characteristics are used: precision, recall, and accuracy – as well as their combinations like F1. In this paper, we use the basic decision theory to explain why these three characteristics are most frequently used.

1 Formulation of the problem

Need to gauge the quality of classification methods. 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.

How this quality is gauged. In principle, we can have many different combinations of these four numbers. Empirically, the following three combinations are most frequently used (see, e.g., [3]):

- *precision*
$$P = \frac{TP}{TP+FP},$$
- *recall*
$$R = \frac{TP}{TP+FN},$$

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, Texas 79968, USA
e-mail: dadavis3@miners.utep.edu, naescobedo@miners.utep.edu, laguevara3@miners.utep.edu, bgrodilesde@miners.utep.edu, arodriguez326@miners.utep.edu, vladik@utep.edu

- and *accuracy*
$$A = \frac{TP+TN}{TP+TN+FP+FN}.$$

Also, the following combination of these characteristics is frequently used:

$$F1 = \frac{P \cdot R}{P+R}.$$

A natural question. Why these characteristics and not others?

What we do in this paper. In this paper, we explain why these characteristics are used.

2 Analysis of the problem

General idea. Each correct classification brings benefits, each false classification brings losses. According to decision theiory, the method should be applied if the benefits are larger than the losses; see, e.g., [1, 2, 4, 5, 6, 7, 8].

Benefits: possible cases. 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:

- the case when $b_{TP} \approx b_{TN}$,
- the case when $b_{TP} \gg b_{TN}$, and
- the case when $b_{TN} \gg b_{TP}$.

Let us simplify the situation. If TN brings more benefits, we can simply rename negative to positive. So, without losing generality, we can say that we have two cases:

- case when $b_{TP} \approx b_{TN}$, and
- case when $b_{TP} \gg b_{TN}$.

In the first approximation, when we ignore small numbers and small differences:

- the first case means $b_{TP} = b_{TN}$, and
- the second case means $b_{TP} > 0$ and $b_{TN} = 0$.

Losses: possible cases. 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 $\ell_{FP} = 0$.

What we plan to do. We have two possible cases for benefits; for each of them, we have three possible cases for losses. Thus, we have $2 \times 3 = 6$ possible situations. Let us start considering these situations one by one.

3 Why precision: an explanation

The case. Let us first consider the case when:

- for benefits, we have $b_{TP} > 0$ and $b_{TN} = 0$, and
- for losses, we have $\ell_{FP} > 0$ and $\ell_{FN} = 0$.

Analysis of this case. In this case, the method is beneficial if $b_{TP} \cdot TP > \ell_{FP} \cdot FP$, i.e., equivalently, when

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

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. For example, we can take a strictly increasing function

$$\frac{1}{1 + 1/r_1}.$$

Applying this function to

$$r_1 = \frac{TP}{FP},$$

we get exactly the precision.

This explains why precision is used.

4 Why recall: an explanation

The case. Let us now consider the case when:

- for benefits, we have $b_{TP} > 0$ and $b_{TN} = 0$, and
- for losses, we have $\ell_{FN} > 0$ and $\ell_{FP} = 0$.

Analysis of this case. In this case, the method is beneficial if $b_{TP} \cdot TP > \ell_{FN} \cdot FN$, i.e., equivalently, when

$$r_2 \stackrel{\text{def}}{=} \frac{TP}{FN} > \frac{\ell_{FN}}{b_{TP}}.$$

The larger the ratio r_2, the more cases when this method is useful. So, the quality of the method is larger if the ratio r_2 is larger.

Alternatively, we can take any strictly increasing function of r_2. For example, we can take a strictly increasing function

$$\frac{1}{1+1/r_2}.$$

Applying this function to

$$r_2 = \frac{TP}{FN},$$

we get exactly the recall.

5 Why accuracy: an explanation

The case. Let us now consider the case when:

- for benefits, we have $b_{TP} = b_{TN}$, and
- for losses, we have $\ell_{FP} = \ell_{FN}$.

Analysis of this case. In this case, the method is beneficial if

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

Since $b_{TP} = b_{TN}$ and $\ell_{FP} = \ell_{FN}$, this inequality is equivalent to

$$b_{TP} \cdot (TP + TN) > \ell_{FP} \cdot (FP + FN),$$

i.e., equivalently, to

$$r_3 \stackrel{\text{def}}{=} \frac{TP+TN}{TP+TN+FP+FN} > \frac{\ell_{FP}}{b_{TP}}.$$

The larger the ratio r_3, the more cases when this method is useful. So, the quality of the method is larger if the ratio r_3 is larger.

Alternatively, we can take any strictly increasing function of r_3. For example, we can take a strictly increasing function

$$\frac{1}{1+1/r_3}.$$

Applying this function to

$$r_3 = \frac{TP + TN}{FP + FN},$$

we get exactly the accuracy.

6 Why only these three?

Natural question. In general, we can have many different characteristics, so why only there three are used – as well as characteristics like F1 that are obtained by combining these three major characteristics?

Our answer to this question. Our answer is that all possible decision-related characteristics can indeed be obtained as a combination of the basic three ones.

Indeed, 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

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

with three unknown ratios

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

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 basic characteristics P, R, and A depend only on these three ratios:

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

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

$$R_1 = \frac{1}{P} - 1, \;\; R_2 = \frac{1}{A} - 1, \text{ and } R_3 = \frac{A \cdot (R_1 + R_2 + 1) - 1}{1 - A}.$$

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 can be.

Acknowledgments

This work was supported in part by the 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), and by the AT&T Fellowship in Information Technology.

It was also supported by a grant from the Hungarian National Research, Development and Innovation Office (NRDI), by the Institute for Risk and Reliability, Leibniz Universitaet Hannover, Germany, and by the European Union under the project ROBOPROX (No. CZ.02.01.01/00/22 008/0004590).

References

1. P. C. Fishburn, *Utility Theory for Decision Making*, John Wiley & Sons Inc., New York, 1969.
2. P. C. Fishburn, *Nonlinear Preference and Utility Theory*, The John Hopkins Press, Baltimore, Maryland, 1988.
3. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, Cambridge, Massachusetts, 2016.
4. V. Kreinovich, "Decision making under interval uncertainty (and beyond)", In: P. Guo and W. Pedrycz (eds.), *Human-Centric Decision-Making Models for Social Sciences*, Springer Verlag, 2014, pp. 163–193.
5. R. D. Luce and R. Raiffa, *Games and Decisions: Introduction and Critical Survey*, Dover, New York, 1989.
6. H. T. Nguyen, O. Kosheleva, and V. Kreinovich, "Decision making beyond Arrow's 'impossibility theorem', with the analysis of effects of collusion and mutual attraction", *International Journal of Intelligent Systems*, 2009, Vol. 24, No. 1, pp. 27–47.
7. H. T. Nguyen, V. Kreinovich, B. Wu, and G. Xiang, *Computing Statistics under Interval and Fuzzy Uncertainty*, Springer Verlag, Berlin, Heidelberg, 2012.
8. H. Raiffa, *Decision Analysis*, McGraw-Hill, Columbus, Ohio, 1997.