How to Gauge the Quality of a Multi-Class Classification When Ground Truth Is Known with Uncertainty

Ricardo Mendez, Osagumwenro Osaretin, and Vladik Kreinovich

Department of Computer Science, University of Texas at El Paso
500 W. University, El Paso, Texas 79968, USA
{ooosaretin,rmendez13}@miners.utep.edu, vladik@utep.edu

1. Formulation of the Problem

- Traditional methods of gauging the quality of a classification method assume that we know the ground truth.
- In other words, we assume that for some elements, we know, with certainty, to which class they belong.
- E.g., in medical diagnostics, we assume that for some patients, we know, with absolute certainty, what was the correct diagnosis.
- In real life, however, we are rarely absolutely certain.
- Usually, there is some degree of uncertainty, some of the "known" classification may turn out to be wrong.

2. Formulation of the Problem (cont-d)

- Because of this:
  - the values $\widetilde{v}$ of the quality measures that we get when we assume the known classifications to be absolutely true
  - are, in general, different from the ideal values v – that we would have gotten if we knew the actual ground truth.
- How can we gauge the resulting uncertainty in v?
- In the previous papers, this problem was analyzed for the case of 2-class ("yes"-"no") classification.
- In this talk, we start extending these ideas and results to the general multi-class case.
- Specifically, we analyze the uncertainty in accuracy.

3. Notations: Traditional Approach

- Traditional methods assume that we know the ground truth.
- C is the number of possible classes.
- Classes will be denoted by numbers $c = 1, 2, \ldots, C$.
- N is the number of objects whose classification we know.
- P_c is the set of all the objects in the c-th class.
- S_c is the set of all objects that the tested method classifies as belonging to the c-th class.
- $|S|$ is the number of elements in the set S.
- Accuracy A is the proportion of correctly classified objects:

$$A = \frac{M}{N}, \text{ where } M \stackrel{\text{def}}{=} \sum_{c=1}^{C} |P_c \cap S_c|.$$

4. Realistic Approach: Formulation of the Problem

- In practice, experts are not 100% sure about their classification.
- We have the number $\widetilde{N}$ of objects about which experts provided opinions.
- We know the sets $\widetilde{P}_c$ of all objects that experts classified to the i-th class.
- For each object i, we know the expert's probability p_i that his/her classification of this object is correct.
- We estimate the accuracy as $\widetilde{A} = \dfrac{\sum\limits_{c=1}^{C} |\widetilde{P}_c \cap S_c|}{\widetilde{N}}$.
- *Challenge:* how close is this estimate to the actual accuracy A?

5. Our Solution

- Let $\xi(i)$ be 0 or 1 depending on whether the expert's classification of the i-th object is correct; then:
  - with probability p_i, we have $\xi(i) = 1$, and
  - with the remaining probability $1 - p_i$, we have $\xi(i) = 0$.
- Thus, the mean value and the variance are

$$E[\xi(i)] = p_i \text{ and } V[\xi(i)] = p_i \cdot (1 - p_i).$$

- In these terms, $A = \dfrac{M}{N}$, where:

$$N = \sum_{i=1}^{\widetilde{N}} \xi(i) \text{ and } M = \sum_{c=1}^{C} |P_c \cap S_c| = \sum_{i \in \bigcup\limits_{c=1}^{C} E_c \cap S_c} \xi(i).$$

6. Our Solution (cont-d)

- For large $\widetilde{N}$, a linear combination of a large number of relatively small independent random variables is, in effect, normally distributed.
- This follows from the Central Limit Theorem.
- Thus, both N and M are normally distributed.
- We can therefore find the distribution of A as the ratio of two random variables M/N with a joint normal distribution.
- A joint normal distribution is uniquely determined by its means, variances, and covariance; here:

$$E[N]=\sum_{i=1}^{\widetilde{N}} p_i,\quad V[N]=\sum_{i=1}^{\widetilde{N}} p_i\cdot(1-p_i),$$

$$E[M]=\sum_{i\in\bigcup\limits_{c=1}^{C} E_c\cap S_c} p_i,\quad V[M]=\sum_{i\in\bigcup\limits_{c=1}^{C} E_c\cap S_c} p_i\cdot(1-p_i).$$

7. Our Solution (cont-d)

- Here, $N-M$ and M contain different variables and are, thus, independent.
- Similarly, $(N-E[N])-(M-E[M])$ and $M-E[M]$ are also independent, with mean 0; thus:

$$E[((N-E[N])-(M-E[M]))\cdot(M-E[M])] =$$

$$E[(N-E[N])-(M-E[M])]\cdot E[(M-E[M])] = 0.$$

- Hence, for the covariance, we get

$$C(N,M)\stackrel{\text{def}}{=}E[(N-E[N])\cdot(M-E[M])] =$$

$$E[((N-E[N])-(M-E[M]))\cdot(M-E[M])]+E[(M-E[M])^2] = V[M].$$

8. References

- N. Gray, S. Ferson, and V. Kreinovich, "How to gauge the quality of a testing method when ground truth is known with uncertainty", *Proceedings of the 9th International Workshop on Reliable Engineering Computing REC'2021*, Taormina, Italy, May 16–20, 2021, pp. 265–278.

9. 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), and
  - HRD-1834620 and HRD-2034030 (CAHSI Includes).
- It was also supported by the AT&T Fellowship in Information Technology.
- It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478.