Tutorial: Advanced Accuracy Assessment for Binary Classifiers

Prof. Dr. Luigi Lavazza
Dipartimento di Scienze teoriche e Applicate - Università degli Studi dell'Insubria, Italy

Summary

Being able to assess the accuracy of binary classifiers is particularly important. Consider for instance diagnostic tests, which are expected to tell whether a person is affected by a given disease, based on some diagnostic data. In these cases, using the most accurate predictor may be of (literally) vital importance. In other cases, e.g., when deciding whether a given software module if likely defective and should undergo additional quality assurance activities, the issue is less critical, but nonetheless quite relevant from an economic point of view: one does not want to waste money on verifying correct modules, while not verifying defective modules results usually in huge postrelease costs.

Traditionally, receiver operating characteristics (ROC) curves and the area under the curve (AUC) have been used as accuracy indicators, often to choose the most accurate of a set of given classifiers. In this tutorial we show that the usage of AUC may be deceiving, and propose a few considerations that let us identify the truly representative parts of the area under the curve. As a result, we shall get indicators that are much more reliable than traditional ones. In addition, the proposed indicators can be tailored to represent the accuracy aspects that the user deems most important.

Examples from the medical and software engineering fields will be shown.