

CLASSEE - a Visualization Tool for Accessible Evaluation of Classification Performance

Emma Beauxis-Aussalet - CWI - emma@cwi.nl

Joost van Doorn - CWI - joost.van.doorn@cwi.nl

Lynda Hardman - Utrecht University, CWI - lynda.hardman@cwi.nl

Max Welling - University of Amsterdam - m.welling@uva.nl

CWI - Science Park 123 - 1098XG Amsterdam - 020 592 9333

Abstract

Machine Learning techniques for automatic classification have reached a broad range of applications. But the technology transfers face issues with user trust and acceptance, as classification results inherently contain errors. Machine Learning experts rely on widely-established error measurement methods and uncertainty visualizations. However end-users are not familiar with these uncertainty visualizations, and underlying error measures. Simplified visualization designs were proposed to address this issue (Fig. 1). Machine Learning experts showed interest in using such designs to communicate with end-users. However, they wish to continue using the expert visualizations they are most familiar with. Hence we developed an interactive interface to explore classification uncertainty using visualization alternatives. We address the needs of technology providers who continuously improve classification algorithms, and communicate their performance for different application domains. We designed simple interactions for navigating through classifiers, datasets and visualizations. Our tool is developed with the D3 library, and the visualization components are delivered as open-source tools.

Keywords— Data Visualization, Classification Uncertainty, Technology Transfer

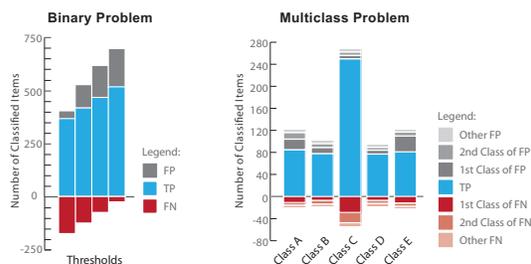


Fig. 1: Visualization of classification errors

1 Use Case

Machine Learning classifiers can analyse the content of a variety of information sources by i) detecting objects of interests (e.g., binary classification); and ii) classify them in descriptive categories (e.g., multi-class classification). Sets of classified items can be affected by two types of errors. Some items may be erroneously added to a category while truly belonging to another (*Type I Error* or *False Positive*), and some items may be missing as they remain undetected or classified in another category (*Type II Error* or *False Negative*). The classification errors can be measured using groundtruth sets of items that are manually classified into the true categories. True classifications are compared to that of classifier outputs, and types I and II errors are encoded in confusion matrices. Machine Learning experts typically use ROC and Precision/Recall curves to visualize these errors. Prior work highlighted that these visualizations are difficult to understand for end-users who are not familiar with the technology [1] and developed simplified alternatives [2, 3] (Fig. 1). Their design choices are to i) show both numbers and rates of errors; ii) omit True Negatives as these are not contained in classifiers' output; iii) use the horizontal axis to separate missed items (below) from selected items (above). Machine Learning researchers were interested in using these to communicate their results to end-users, but also wished to keep ROC and Precision/Recall curves to communicate with other experts. They also needed to compare algorithms' performance over different groundtruth sets used for training or testing them.

2 Prototype

We designed an interface to i) seamlessly use alternative visualizations; ii) explore classification results for different algorithms and groundtruth sets (Fig. 2). Machine learning experts can explore clas-

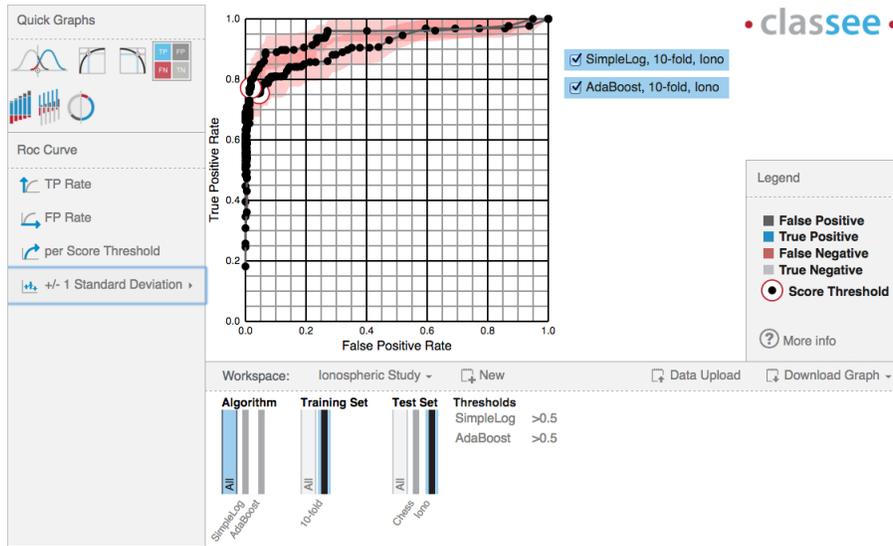


Fig. 2: Screenshot of the visualization tool

sification uncertainty by selecting the algorithms, groundtruth sets and classes of interest. These can be selected using dedicated widgets at the bottom of the interface. Classification errors are visualised using interchangeable graphs at the top of the interface. These can be selected using the left panel. The top icons provide 1-click access to most common graphs (e.g., ROC and Precision/Recall curves, confusion matrix table, and the simplified visualizations in Fig. 1). Below are options to fine tune the visualization (e.g., to display estimates of error rate variance). Machine Learning experts can experiment with algorithms and groundtruth sets using expert visualizations. Once they want to communicate their results to end-users, they can seamlessly switch the top visualization to simplified graphs, and disseminate them by using the download option.

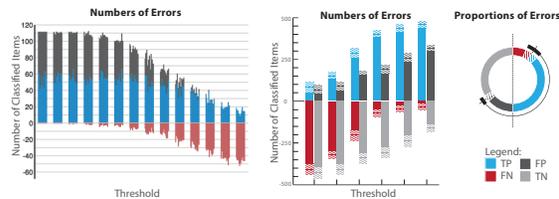


Fig. 3: Visualizations of error variance

3 Advanced Features

Our tool can show the potential variance of classification errors by using estimates from sampling theory (Fig. 2). The stacked chart design in Fig. 1 is not appropriate to show standard deviations, which in

theory cannot always be summed or stacked. Hence we also integrated simplified user-friendly visualizations for exploring error rate variability (Fig. 3), without using potentially misleading error bars [5].

References

- [1] Beauxis-Aussalet, E., et al.: A case study of trust issues in scientific video collections. Proceedings of the 2nd ACM International Workshop on Multimedia Analysis for Ecological Data (MAED), 2013
- [2] Beauxis-Aussalet, E., Hardman, L.: Simplifying the Visualization of Confusion Matrix. Proceedings of the Benelux Conference on Artificial Intelligence (BNAIC), 2014.
- [3] Beauxis-Aussalet, E., Hardman, L.: Visualization of Confusion Matrix for Non-Expert Users. Poster at the IEEE Symposium on Information Visualization (IEEE Vis), 2014.
- [4] Beauxis-Aussalet, E., van Doorn, J., Hardman, L.: Bridging the Gap Between Machine Learning Experts and End-Users with Interactive Uncertainty Visualization. Poster at the Eurographics Conference on Visualization (EuroVis), 2015.
- [5] Correll, M., Gleicher, M.: Error Bars Considered Harmful: Exploring Alternate Encodings for Mean and Error. IEEE Transactions on Visualization and Computer Graphics, 20(12) pp.2142-2151, 2014.