Eurographics Conference on Visualization (EuroVis) (2015)
R. Maciejewski and F. Marton (Editors) *Posters*

# Bridging the Gap Between Machine Learning Experts and End-Users with Interactive Uncertainty Visualization

Emma Beauxis-Aussalet, Joost van Doorn and Lynda Hardman

CWI, The Netherlands

**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. 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 the 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 describe the interaction design for navigating through datasets and visualizations. The interactive prototype is developed with D3 library and the visualization components will be delivered as open-source tools. We conclude by discussing future developments of the interface features.*

Categories and Subject Descriptors (according to ACM CCS): Information Interface and Presentation (e.g., HCI) [H.5.2]: Prototyping—Artificial Intelligence [I.2.1]: Applications and Expert Systems—

## 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., multiclass classification). Sets of classified items can be affected by two types of errors.

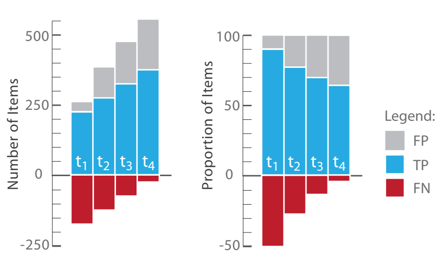


**Figure 1:** *Simplified visualization for binary classification.*

Some items may be erroneously added to the 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 (Fig. 2). Machine Learning experts typically use ROC and Precision/Recall curves to visualize these errors [Faw06]. Prior work highlighted that theses visualizations are difficult to understand for end-users who are not familiar with the technology [BAAHVO13] and developed simplified alternatives [BAH14] (Fig. 1,3). 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). A team of Machine Learning researchers was interested in using these to communicate their results to end-users, but also wished to

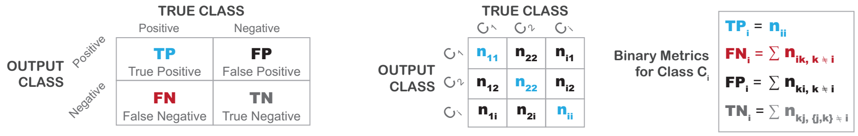


**Figure 2:** *Confusion matrices for binary (left) and multiclass (right) classification tasks.*

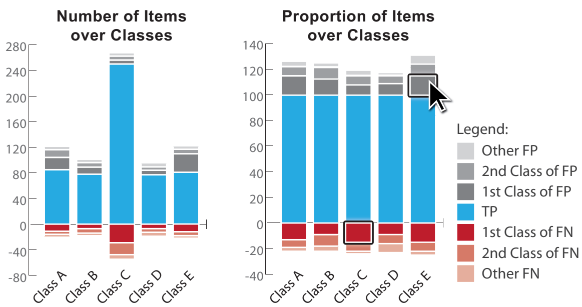


**Figure 3:** *Visualization for multiclass classification.*

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 groundtruth sets (Fig. 4). Machine learning experts can explore classification uncertainty by selecting the algorithms, groundtruth sets and classes of interest. These can be selected using dedicated widgets at the bottom of the interface. The histograms show overviews of numbers of groundtruth items. Classification errors are visualised using interchangeable graphs at the top of the interface. These are 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. 1,3). Below are options to fine tune the visualization (e.g., to display other error rates such as F-measures). 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 or screenshots.

## 3. Discussion

Additional interface features may be of interest for experts or end-users. Experts wish to visualize the variance of results over different groundtruth sets. But the stacked chart design in Fig. 1,3 is not appropriate to show, e.g., standard

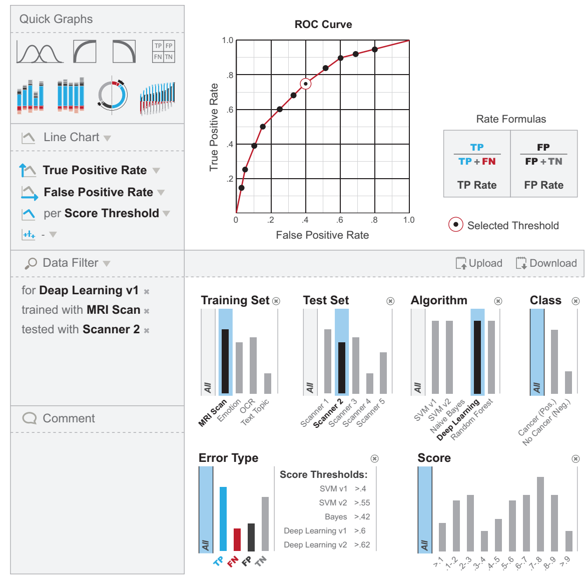


**Figure 4:** *Prototype of the interactive visualization.*

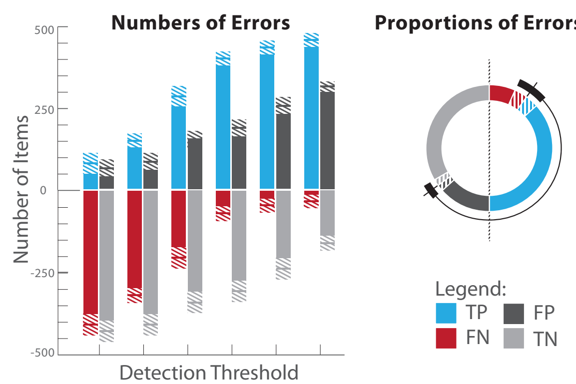


**Figure 5:** *Alternative visualization with error variance.*

deviations (which in theory cannot always be additioned). Hence we devised alternative visualizations (Fig. 5) without error bars [CG14]. Finally, end-users may wish to explore the datasets and algorithms to make their own decisions. A simplified set of options may be devised to omit, e.g., complex error rates such as F-measure. To decide on the appropriate set of features, we will empirically evaluate our interactive prototype and individual visualizations with both technology and domain experts.

## References

[BAAHVO13] BEAUXIS-AUSSALET E., ARSLANOVA E., HARDMAN L., VAN OSSENBRUGGEN J.: A case study of trust issues in scientific video collections. In *Proceedings of the 2nd ACM international workshop on Multimedia analysis for ecological data* (2013), ACM, pp. 41–46. 1

[BAH14] BEAUXIS-AUSSALET E., HARDMAN L.: Simplifying the visualization of confusion matrix. In *BNAIC* (2014), pp. 132–133. 1

[CG14] CORRELL M., GLEICHER M.: Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (2014), 2142–2151. 2

[Faw06] FAWCETT T.: An introduction to ROC analysis. *Pattern recognition letters 27*, 8 (2006), 861–874. 1