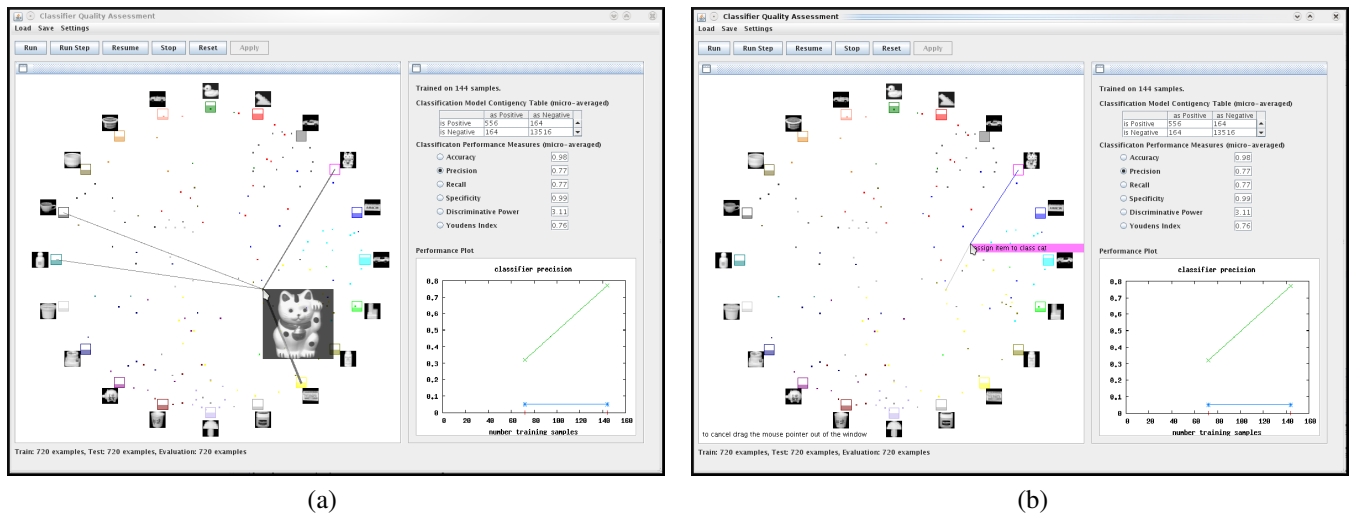# A Visualization to Investigate and Give Feedback to Classifiers

Christin Seifert and Elisabeth Lex
Know-Center Graz, Austria*

Figure 1: *(a) Selection of an item shows the classifier decision. The thickest line and the color of the item point indicated that the item is false classified. (b) Moving the item towards the correct class retrains the classifier.*

1 Introduction

Classification is a core technique of data mining. The performance of a classifier might be improved by user feedback especially when a stable classifier cannot be generated from the available training data. Classification processes and classifier decisions are sometimes hard to grasp. However, for giving feedback users need to understand a classifier's decision and the results. Visualizations can boost the understanding of classification processes and results even for non-expert users. In the past, several classifier visualizations were proposed, however they are either restricted to a limited set of classifiers [Chodos and Zaiane 2008] or no feedback mechanisms are supported [Rheingans and desJardins 2000].

In this work we present a visualization system to investigate and give feedback to all classifiers whose predictions can be interpreted as probability distribution over classes. We provide an intuitive and simple application to observe classification processes and results even for non-expert users. Users can give feedback to classifiers directly in the visualization with well-known interaction techniques like drag and drop.

2 Our Approach

With our system users can investigate classification processes on three levels, on classifier level, on class level and on test item level. The classifier level provides a visual overview of the classifier performance. All classes are uniquely colored and the test items are drawn in the color of the assigned class. Additionally, common performance measures are computed and displayed. On class level, a mouse over shows general information and performance measures for a specific class. On test item level, the position of the item within the visualization indicates how well it could be classified. Besides, the a-posteriori probability values above a threshold are displayed by lines while line thickness denotes the probability values. Via mouse over, the underlying item data is provided. Figure 1 shows an example of the visualization. We performed an image recognition task with 20 classes on the COIL image database[1] using a k-NN classifier. In Figure 1(a) a test item obviously assigned to the wrong class is selected. The selected item is assigned to the class "tylenol" (indicated by the thickest a-posteriori line) whereas the test item actually belongs to the class "cat" (shown by the underlying image data). The user now can move the item towards the correct class "cat" as shown in Figure 1(b). The moved item then serves as new training example for the particular class and the classifier is re-trained and re-evaluated.

We will conduct a user study to find out whether the visualization and the feedback mechanisms are really intuitive and clear to users. Further the ambiguity problem - more than one item can be mapped to the same location - needs to be tackled.

Acknowledgements

References

CHODOS, D., AND ZAIANE, O. 2008. ARC-UI: Visualization Tool for Associative Classifiers. In *Proc. International Conference on Information Visualisation*, 296–301.

RHEINGANS, P., AND DESJARDINS, M. 2000. Visualizing high-dimensional predictive model quality. In *In Proceedings of IEEE Visualization*, 493–496.

---

*e-mail:{cseifert,elex}@know-center.at

[1]http://www.cs.columbia.edu/CAVE//software/softlib/coil-20.php