

Incremental Exemplar Learning Schemes for Classification on Embedded Devices*

Ankur Jain and Daniel Nikovski

Mitsubishi Electric Research Laboratory, 201 Broadway, Cambridge, MA 02139
{jain,nikovski}@merl.com

In this paper, we focus on the data classification problem when the classifier operates on an embedded device (e.g., fault detection in device condition-monitoring data streams). Memory-based classifiers are an excellent choice in such cases, however, an embedded device is unlikely to be able to hold a large training dataset in memory (which could potentially keep increasing in size as new training data with new concepts arrive). A viable option then is to employ exemplar learning (EL) techniques to find a training subset comprising a few carefully selected *exemplars* of high functional value that fit in memory and effectively delineate the class boundaries. We propose two novel incremental EL schemes that unlike traditional EL approaches [3] are, (1) incremental (they naturally incorporate new training data streams), (2) offer ordered removal of instances (they can be customized to obtain exemplar sets of any user-defined size) and (3) robust (such that the exemplar sets generalize for other classifiers as well). Our proposed methods are as follows:

- EBEL (Entropy Based EL) – This method removes instances from the training set based on their *information content*. Instead of using an adhoc ranking scheme, it removes a training instance whose removal causes the least amount of drop in the conditional entropy of the class indicator variable insuring minimum loss of information.
- ABEL (AUC Based EL) – This method prunes data based on AUC (Area under ROC curve) performance. ABEL uses a *validation set* and prunes an instance if its removal offers the least drop in the AUC computed for this validation set.

We show that our schemes efficiently incorporate new training datasets while maintaining high-quality exemplar sets of any user-defined size. We present a comprehensive experimental analysis showing excellent classification-accuracy versus memory-usage tradeoffs of our proposed methods.

References

1. Aha, D.W., Kibler, D., Albert, M.K.: Instance-based learning algorithms. *Mach. Learn.* 6(1), 37–66 (1991)
2. Jain, A., Nikovski, D.: Incremental exemplar learning schemes for classification on embedded devices. *Machine Learning* 72(3), 189–203 (September 2008)
3. Wilson, D.R., Martinez, T.R.: Reduction techniques for instance-based learning algorithms. *Machine Learning* 38(3), 257–286 (2000)

* This is an extended abstract of an article published in the *Machine Learning* journal [2].