Development and experimental evaluation of an adaptive data stream classifier

Luke Marsden

BSc in Computer Science with Industrial Year

ABSTRACT

Data stream mining has a greater set of problems in comparison to data set mining. Data stream classifiers need the ability to re-learn and evolve their decision models on the fly, unable to rely on the assumption of having a fixed attribute distribution within future instances. Evaluating the performance of these classifiers also brings problems, and researching and implementing these evaluation techniques is the focus of this study. Various evaluation metrics and techniques are discussed, alongside the new behaviour exhibited by the eRules and G-eRules – the ability to abstain from classifying an unseen instance. This behaviour allows additional performance metrics to be explored within evaluations – no existing tool does this. The resulting product from this study is an application which performs various data stream classifier evaluations and display detailed results graphically to the user. This tool can be used to evaluate any classifier present in the Massive Online Analysis framework.

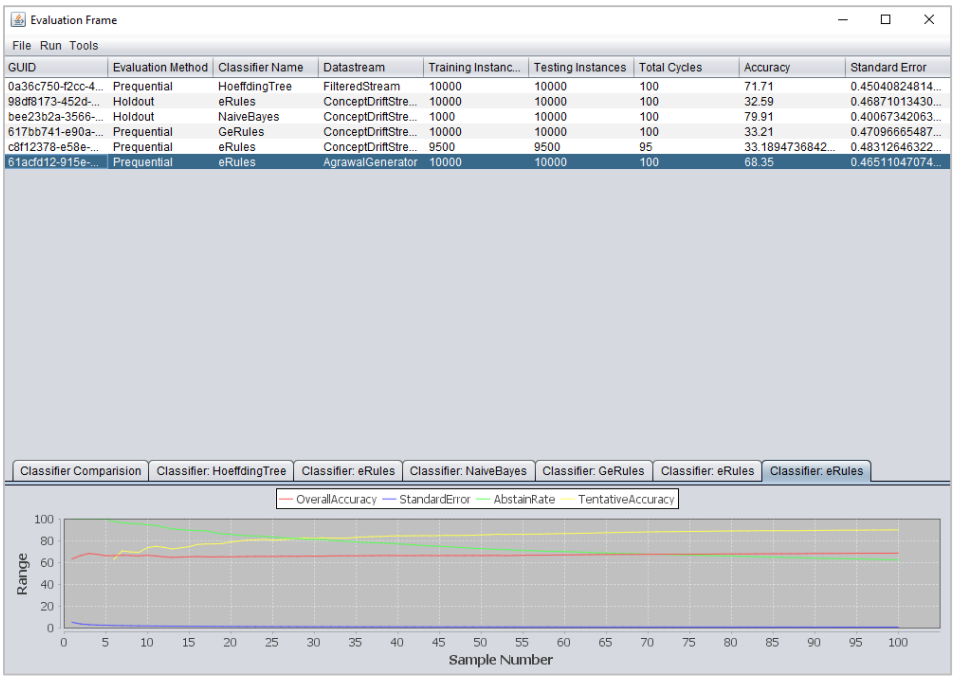

Figure 1. Main application interface displaying multiple evaluations and graphical results

Luke Marsden, Development and experimental evaluation of an adaptive data stream classifier, *Proc. 13th School Conf. for Annual Research Projects*, V F Ruiz (Ed), pp. xx–yy, University of Reading, 24th May 2016.