

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.

	ne							- 🗆 ×
File Run Tools								
GUID	Evaluation Method	Classifier Name	Datastream	Training Instanc	Testing Instances	Total Cycles	Accuracy	Standard Error
0a36c750-f2cc-4	Prequential	HoeffdingTree	FilteredStream	10000	10000	100	71.71	0.45040824814
98df8173-452d	Holdout	eRules	ConceptDriftStre	10000	10000	100	32.59	0.46871013430
bee23b2a-3566	Holdout	NaiveBayes	ConceptDriftStre	1000	10000	100	79.91	0.40067342063
617bb741-e90a	Prequential	GeRules	ConceptDriftStre	10000	10000	100	33.21	0.47096665487
c8f12378-e58e	Prequential	eRules	ConceptDriftStre	9500	9500	95	33.1894736842	0.48312646322
61acfd12-915e	Prequential	eRules	AgrawalGenerator	10000	10000	100	68.35	0.46511047074
					^			
Classifier Comn	arision Classifier I	HoeffdingTree C	lassifier eRules Cl	assifier: NaiveBaves	Classifier GeRul	es I Classifier el	Rules Classifier eF	lules
Classifier Comp	arision Classifier: I	HoeffdingTree C	lassifier: eRules Cl	assifier: NaiveBayes	Classifier: GeRul	es Classifier: el	Rules Classifier: eF	tules

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.