Label Free Change Detection on Streaming Data with Cooperative Multi-objective Genetic Programming

Sara Rahimi
Faculty of Computer Science
Dalhousie University
Halifax, Canada, B3H 4R2
srahimi@cs.dal.ca

Andrew R. McIntyre
Faculty of Computer Science
Dalhousie University
Halifax, Canada, B3H 4R2
armcnty@cs.dal.ca

Malcolm I. Heywood
Faculty of Computer Science
Dalhousie University
Halifax, Canada, B3H 4R2
mheywood@cs.dal.ca

Nur Zincir-Heywood
Faculty of Computer Science
Dalhousie University
Halifax, Canada, B3H 4R2
zincir@cs.dal.ca

ABSTRACT

Classification under streaming data conditions requires that the machine learning (ML) approach operate interactively with the stream content. Thus, given some initial ML classification capability, it is not possible to assume that stream content will be stationary. It is therefore necessary to first detect when the stream content changes. Only after detecting a change, can classifier retraining be triggered. Current methods for change detection tend to assume an entropy *filter* approach, where class labels are necessary. In practice, labeling the stream would be extremely expensive. This work proposes an approach in which the behaviour of GP individuals is used to detect change *without* the use of labels. Only after detecting a change is label information requested. Benchmarking under a computer network traffic analysis scenario demonstrates that the proposed approach performs at least as well as the filter method, while retaining the advantage of requiring no labels.

Categories and Subject Descriptors

I.2.6 [**Learning**]: Parameter Learning

General Terms

Algorithms, Experimentation, Performance

Keywords

Streaming data, Dynamic environments, Coevolution, Pareto archiving, Genetic Programming

1. INTRODUCTION

Streaming data represents a distinct set of requirements from that classically assumed by 'batch' or off-line classification. For example, off-line classification assumes that independent training and test sets exist and that the content of the training set can be revisited without penalty. In the case of genetic programming (GP) a population of candidate classifiers would be evolved against the content of the training set and post training evaluation is performed relative to the test set. Various monographs have discussed evolutionary computation as applied under non-stationary or 'dynamic' environments[1]. We are mainly interested in decoupling the need for supplying class labels at the rate of the stream. In particular, we focus on the concept of change detection. Thus rather than attempt to perform evolution on a continuous basis we are first interested in detecting when a change occurs. If we can do this without reference to labels, then we will only trigger retraining under specific circumstances. Two generic approaches have been identified to date, either an entropy based measure as applied to the input stream (e.g., [2]) or through a behavioural analysis of a team (or ensemble) of classifiers (e.g., [3]). In this work, we will consider both scenarios, the former providing the performance baseline that the latter should ideally retain. From the perspective of applications in network traffic analysis, this is a very important requirement because the cost of providing labels is high.

2. STREAMING CMGP

CMGP represents a framework for GP in which multiple (GP) individuals are evolved to cooperatively represent each class of a classification task [1]. There are three components that make it useful to the streaming task, specifically Pareto archiving (decouples the cost of fitness evaluation from the cardinality of the training set (TS)); gaussian local membership function (LMF, provides support for problem decomposition); and evolutionary multi-objective optimization (EMO) as the basis for credit assignment. See [1] for CMGP details. From a streaming data perspective, we are interested in characterizing changes in data from the stream without use of labels by measuring the degree of variation relative to CMGP's learner archive (LA) content. To begin, the initial data content of TS is used to construct the LA content and a sliding window source into the data stream is established. LA individuals provide labels for each exemplar under the current sliding window location. The result of executing each GP champion for each exemplar is a degree of

[1]Hereafter non-stationary and dynamic will be employed interchangeably.

GECCO'13 Companion, July 6–10, 2013, Amsterdam, The Netherlands.
ACM 978-1-4503-1964-5/13/07.

LMF membership associated with its mapping onto *gpOut*. Any change in the ratio of class labels between PA and the sliding window is not necessarily indicative of a change in the underlying process describing the stream data; however, if a shift takes place in the process describing the data, we might expect this to be reflected in the confidence (LMF values) associated with the winning LA champion. Thus, the LMF values for PA is recorded and compared to that of the current sliding window. A student T-test is applied to each and only when a significant difference is registered do we update the content of the point population (PP) and perform a fixed number of training epochs (25) to update LA and PA relative to the new points. Hereafter this algorithm is denoted Behavioural CMGP or **Bv-GP**.

3. RESULTS

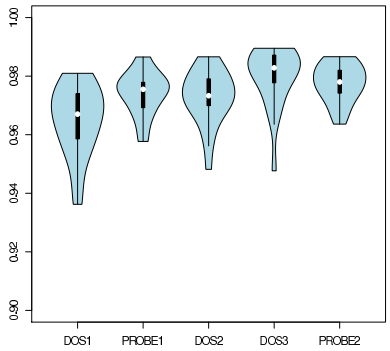
Benchmarking is reported under a network traffic analysis application. The proposed scheme for unlabelled change detection is shown to at least match that of those requiring all exemplars to be labeled. In all cases, we assume that some initial labeled sample of data is available from the stream and used to seed TS and all CMGP configurations share a common parameterization. Three variations will then be considered: (1) **Baseline CMGP** algorithm, i.e., no change detection, no additional training; (2) **Entropy filter** uses classical label-based change detection as applied to sliding window pairs from the data stream where CMGP labels the stream using the current LA content, with retraining for a fixed number of cycles (i.e., updates to LA) being instigated by change detection care of the entropy filter (entropy based T-test); (3) **Behavioural CMGP** we characterize the difference between the 'confidence' of CMGP's LA champions actually supplying the labels between a pair of sliding windows (entirely independent of label information). Only when a significant difference appears in the confidence between pairs of windows are labels requested and then only for the window triggering the event. We discuss results of the baseline and two streaming approaches under a data set constructed from network flow data. In the case of this work up to 40 flow attributes are available care of the open source tool "NETMATE"[2]. Netmate is applied to the KDD'99 contest data set[3]. We are specifically interested in distinguishing between Denial of Service (DoS) and Probing behaviours from 'normal' behaviour. The (post training) stream itself is composed from five distinct blocks and follows a sequence in which blocks representing the abnormal class label appear in the order: DoS-1, Probe-1, DoS-2, DoS-3, Probe-2. All DoS (Probe) blocks consist of 3,000 (4,000) flows.

Evaluation takes the form of average class-wise detection rate (DR) as measured on each block of the stream. In the baseline scenario, CMGP is trained on the initial TS content in batch mode and LA individuals label the stream. Evaluation of the 'streaming enabled' algorithms implies that whenever a significant change is detected[4] the point population is updated and retraining is triggered. The detection rates for the two online algorithms (Figure 1), are statistically independent from the (offline) baseline method across all blocks, illustrating the benefit of change detection and

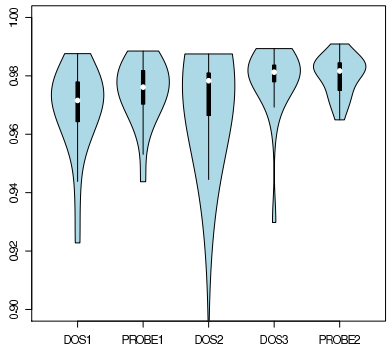
[2]http://dan.arndt.ca/nims/calculating-flow-statistics-using-netmate/

[3]http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[4]Both streaming algorithms use a T-test, p-value 0.9999.



(a) Ent filter



(b) Bv-GP

Figure 1: Average detection rates over KDD stream following pre-training.

retraining. We also observed that Bv-GP did not trigger retraining on the 'Probe-1' block in 50% of the runs. Moreover, there appears to be an incremental decrease in the instances of retraining as the number of blocks increases. We demonstrate that by concentrating on the *behaviour* of the team of pre-trained champion GP individuals, bv-GP not only matches the detection rates of the classical approach to change detection, but does so *without* reference to any label information. Future research will extend the algorithm to the case of gradual changes in stream content.

Acknowledgements

The authors gratefully acknowledge NSERC funding.

4. REFERENCES

[1] A. R. McIntyre and M. I. Heywood. Classification as clustering: A pareto cooperative-competitive gp approach. *Evolutionary Computation*, 19(1):137–166, Mar. 2011.

[2] P. Vorburger and A. Bernstein. Entropy-based concept shift detection. In *Proceedings of the Sixth International Conference on Data Mining*, ICDM '06, pages 1113–1118, Washington, DC, USA, 2006. IEEE Computer Society.

[3] X. Zhu, P. Zhang, X. Lin, and Y. Shi. Active learning from stream data using optimal weight classifier ensemble. *IEEE Transactions on Systems, Man, and Cybernetics–Part B*, 40(6):1607–1621, 2010.