Scalable and efficient multi-label classification for evolving data streams

**Jesse Read · Albert Bifet · Geoff Holmes ·
Bernhard Pfahringer**

Abstract Many challenging real world problems involve multi-label data streams. Efficient methods exist for multi-label classification in non-streaming scenarios. However, learning in evolving streaming scenarios is more challenging, as classifiers must be able to deal with huge numbers of examples and to adapt to change using limited time and memory while being ready to predict at any point.

This paper proposes a new experimental framework for learning and evaluating on multi-label data streams, and uses it to study the performance of various methods. From this study, we develop a multi-label Hoeffding tree with multi-label classifiers at the leaves. We show empirically that this method is well suited to this challenging task. Using our new framework, which allows us to generate realistic multi-label data streams with concept drift (as well as real data), we compare with a selection of baseline methods, as well as new learning methods from the literature, and show that our Hoeffding tree method achieves fast and more accurate performance.

Keywords Multi-label classification · Data streams classification

J. Read currently at: Universidad Carlos III, Spain: E-mail: jesse@tsc.uc3m.es.
A. Bifet currently at: Yahoo! Research Barcelona, Catalonia: E-mail: abifet@yahoo-inc.com.

J. Read · A. Bifet (✉) · G. Holmes · B. Pfahringer
Computer Science Department, University of Waikato, Hamilton, New Zealand
e-mail: abifet@cs.waikato.ac.nz

J. Read
e-mail: jmr30@cs.waikato.ac.nz

G. Holmes
e-mail: geoff@cs.waikato.ac.nz

B. Pfahringer
e-mail: bernhard@cs.waikato.ac.nz