© 2013 FIZ Karlsruhe

io-port 06022931

Ikonomovska, Elena; Gama, João; Džeroski, Sašo

Learning model trees from evolving data streams.

Data Min. Knowl. Discov. 23, No. 1, 128-168 (2011).

Summary: The problem of real-time extraction of meaningful patterns from time-changing data streams is of increasing importance for the machine learning and data mining communities. Regression in time-changing data streams is a relatively unexplored topic, despite the apparent applications. This paper proposes an efficient and incremental stream mining algorithm which is able to learn regression and model trees from possibly unbounded, high-speed and time-changing data streams. The algorithm is evaluated extensively in a variety of settings involving artificial and real data. To the best of our knowledge there is no other general purpose algorithm for incremental learning regression/model trees able to perform explicit change detection and informed adaptation. The algorithm performs online and in real-time, observes each example only once at the speed of arrival, and maintains at any-time a ready-to-use model tree. The tree leaves contain linear models induced online from the examples assigned to them, a process with low complexity. The algorithm has mechanisms for drift detection and model adaptation, which enable it to maintain accurate and updated regression models at any time. The drift detection mechanism exploits the structure of the tree in the process of local change detection. As a response to local drift, the algorithm is able to update the tree structure only locally. This approach improves the any-time performance and greatly reduces the costs of adaptation.

Keywords: non-stationary data streams; stream data mining; regression trees; model trees; incremental algorithms; on-line learning; concept drift; on-line change detection
doi:10.1007/s10618-010-0201-y