

Short introduction to stream analysis using MOA

Massive Online Analysis

Martin Juřen

Fakulta informatiky
Masarykova univerzita

Brno, 2011

Content

- 1 Data stream evaluation
- 2 Work with MOA
 - Basic steps
 - Data stream generators
 - Data stream classifiers
 - Data stream clustering
 - Data stream algorithm evaluation
- 3 MOA Demonstration
- 4 Bibliography

Requirements on algorithms

- Process an example at a time, inspect it only once
- Limited amount of memory and time
- Predict at any time

Basic steps using MOA

- 1 Choose and configure data stream generator
- 2 Choose and configure an algorithm
- 3 Choose and configure an evaluation method

Generators

ArffFileStream Input from file

ConceptDriftStream It generates stream with concept drift

FilteredStream:AddNoiseFilter It generates stream with noise

AgrawalGenerator Based on text: Rakesh Agrawal, Tomasz Imielinski, Arun Swami: Database Mining: A Performance Perspective. IEEE Transactions on Knowledge and Data Engineering, 1993.

Some other generators And their concept drift variants

Classifiers

Majority class most frequently observed class

Hoeffding tree (and variants) This algorithm stands on the fact that small sample could be enough to find an optimal splitting attribute.

Naive Bayes

Decision Stump single-level decision trees

Clustering

CobWeb Not quite data stream algorithm

CluStream Temporal extensions of cluster feature vector, micro-clusters. Storing in snapshots in pyramidal pattern.

ClusTree Parameter-free algorithm. Capable to detect concept drift.

Evaluation

Holdout Periodically testing the model with one test set

Test-Then-Train The model is firstly tested by new data until then it is trained.

Demonstration

- It is a new software, there is a lot of bugs.
- Written in Java, Open source project
- It could be linked to WEKA
- Goal: running experiments, evaluating algorithms, algorithm comparison

Bibliography

- MOA team: moa.cs.waikato.ac.nz
- Biffet A., Kirkby R., Krannen P., Reutemann P.: Massive Online Analysis: Manual. Online at moa.cs.waikato.ac.nz, May 2011.
- Biffet A., Kirkby R.: Tutorial 1. Introduction to MOA: {M}assive {O}nline {A}nalysis. Online at moa.cs.waikato.ac.nz, January 2011.