A Java Toolbox for Analysis of Massive Data Streams using Probabilistic Graphical Models

Masegosa, Andres; Martinez, Ana Maria; Ramos-López, Darío ; Langseth, Helge; Nielsen, Thomas Dyhre; Salmerón, Antonio; Cabanas, Rafael; Madsen, Anders Læsø

Publication date: 2016

A Java Toolbox for Analysis of Massive Data Streams using Probabilistic Graphical Models

Andrés R. Masegosa¹, Ana M. Martínez², Darío Ramos-López³, Helge Langseth¹, Thomas D. Nielsen², Antonio Salmerón³, Rafael Cabañas² & Anders L. Madsen²,⁴

¹ Department of Computer and Information Science, NTNU, Norway ² Department of Computer Science, Aalborg University, Denmark ³ Department of Mathematics, University of Almería, Spain ⁴ Hugin Expert A/S, Aalborg, Denmark

Presentation

Data mining frameworks

- PGMs
- AMIDST
- Stationary data sets
- Data streams

Data streams

- MLlib|Apache Spark/Flink
- MOA
- Elvira
- Infer.net
- Hugin
- Weka
- R Libs
- Matlab
- Apache SAMOA
- Vowpal Wabbit

STATIC

- Naive Bayes
- TAN
- (G)AODE/HODE
- Maximum Likelihood

DYNAMIC

- Dynamic NB
- Gaussian Discriminant Analysis
- Latent Classification Models (LCM)
- Gaussian Mixture
- Bayesian Linear Regression
- Factor Analysis
- Mixture of FA

Bayesian Linear Regression

Use-case: Risk prediction in credit operations

Concept drift

Correlated with Unemployment Rate

And much more...

amidst.eu
amidst.github.io/toolbox/

Code example

```
// We create a 256 bit graph
STreams using Probabilistic Graphical Models

System.out.println(bnModel.toString());
// We print the model

bnModel = parameterLearningAlgorithm.getLearntBayesianNetwork();
// We get the model

parameterLearningAlgorithm.setDataStream(data);
// We set the data which is going to be used for learning the parameters

parameterLearningAlgorithm.setWindowsSize(100);
// We fix the size of the window

parameterLearningAlgorithm.setDAG(DAGGenerator.getHiddenNaiveBayesStructure(data."
// We fix the DAG structure

SVB parameterLearningAlgorithm = new SVB();
// We create a SVB object

DataStream<DataInstance> data = DataStreamLoader.openFromFile("datasets/simulated/
// We can open the data stream using the static class DataStreamLoader

WasteIncineratorSample.arff");
// WasteIncineratorSample.arff is a file with data from a waste incinerator

```

Main Features

- Java 8 based
- Latent variable models
- Integration

Big Data

Open source

- Concept drift
- Correlated with Unemployment Rate

Code example

```
// We create a 256 bit graph
// We can activate the output

```

Use-case: Risk prediction in credit operations

Concept drift

Correlated with Unemployment Rate

And much more...

amidst.eu
amidst.github.io/toolbox/

AMIDST project has received funding from the European Union’s Seventh Framework Programme for research, technological development and demonstration under grant agreement no 619209.