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

Link to publication from Aalborg University

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

Andrés R. Masegosa¹, Ana M. Martínez², Darío Ramos-Lopez³, Helge Langseth¹, Thomas D. Nielsen², Antonio Salmerón³, Rafael Cabañas² & Anders L. Madsen²,³

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

- **Academic and Industrial partners**
  - NTNU
  - Universidade de Almería
  - Daimler

**Description**

- **Analysis of big data streams**: A complete collection of algorithms for inference and learning of both static and dynamic Bayesian networks from streaming data. Existing software systems for PGMs only focus on stationary datasets.
- **Distributed parallel algorithms**: AMIDST provides parallel multi-core and distributed implementations of Bayesian parameter learning, using streaming variational Bayes and variational message passing.

**Main Features**

- Java 8 based
- Latent variable models
- Integration
- Big Data
- Modularity
- Open source

**Code example**

```java
// We create a DMM object
SVMParameterLearningAlgorithm = new SVM();
// We set the data which is going to be used for learning
SVMParameterLearningAlgorithm.setData(data);
// We perform the learning
SVMParameterLearningAlgorithm.runLearning();
// We print the model
System.out.println(bnModel.toString());
```

**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.**