

Analysis of massive data streams using R and AMIDST

Anders L. Madsen[†], Antonio Salmeron[‡]

[†]*Hugin Expert A/S, Aalborg, Denmark*

[‡]*Department of Mathematics, University of Almeria, Spain*

<http://www.hugin.com>

Abstract: Today, omnipresent sensors are continuously providing streaming data on the environments in which they operate. Sources of streaming data with even a modest updating frequency can produce extremely large volumes of data, thereby making efficient and accurate data analysis and prediction difficult. Probabilistic graphical models (PGMs) provide a well-founded and principled approach for performing inference and belief updating in complex domains endowed with uncertainty. The on-going EU-FP7 research project AMIDST (Analysis of MassIve Data SStreams, <http://www.amidst.eu>) is aimed at producing scalable methods able to handle massive data streams based on Bayesian networks technology. All of the developed methods are available through the AMIDST toolbox, implemented in Java 8. We show how the functionality of the AMIDST toolbox can be accessed from R. Available AMIDST objects include variables, distributions and Bayesian networks, as well as those devoted to inference and learning. The interaction between both platforms relies on the rJava package.

Keywords: HUGIN, AMIDST, data streams