Project no.: 619209

Project full title: Analysis of Massive Data STreams

Project Acronym: AMIDST

Deliverable no.: D3.2

Title of the deliverable: Progress report on software development

Contractual Date of Delivery to the CEC:	**31.03.2015**
Actual Date of Delivery to the CEC:	**31.03.2015**
Organisation name of lead contractor for this deliverable:	**AAU**
Author(s):	**Hanen Borchani, Antonio Fernández, Helge Langseth, Anders L. Madsen, Ana M. Martínez, Andrés Masegosa, Thomas D. Nielsen, Antonio Salmerón**
Participants(s):	**P01, P02, P03, P04, P05, P06, P07**
Work package contributing to the deliverable:	**WP2**
Nature:	**R**
Version:	**1.0**
Total number of pages:	**...**
Start date of project:	**1st January 2014 Duration: 36 month**

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)		
Dissemination Level		
PU	Public	**X**
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Abstract:

In this document, we describe the status of the software development in respect to inference.

Keyword list: AMIDST software development, progress report, inference engine.

Contents

Document history

Version	Date	Author (Unit)	Description
v0.3	10/03/2015	All consortium members	The software library implementation for the AMIDST modelling framework discussed and established
v0.6	20/03/2015	Hanen Borchani, Antonio Fernández, Ana M. Martínez, Andrés Masegosa	Initial version of document finished and reviewed
v1.0	31/03/2015	Hanen Borchani, Antonio Fernández, Helge Langseth, Anders L. Madsen, Ana M. Martínez, Andrés Masegosa, Thomas D. Nielsen, Antonio Salmerón	Final version of document

1 Executive summary

2 Introduction

The aim of this document is to provide

A thorough and up-to-date description of the-state-of-the art on inference in hybrid Bayesian networks

3 Preliminaries

Deliverable 3.1 State of the art of inference in hybrid and dynamic models [1].

Variational inference [2, 3]

conjugate-exponential families [3, 6, 9]

An overview of the data structures implemented in the AMIDST toolbox is illustrated in Figure 3.1. These data structures basically define the main components that will be used for implementing the AMIDST learning and inference algorithms. As we previously mentioned, in the AMIDST toolbox, we focus on two specific instantiations of PGMs, namely, a static Bayesian network (BN component) and a two time-slice dynamic Bayesian network (2T-DBN component). This is also directly reflected in the component structure.

4 Inference engine

In this section we provide an overview of the current implementation status of the AMIDST toolbox related to inference algorithms. Figure 4.1 illustrates the main core components of the toolbox. The color coding in the figure summarizes the implementation status: blue bodes represent software components that have been implemented in the AMIDST toolbox and green boxes represent components that are part of the software design specification but which have not yet been implemented.
Note that some of the core components of the AMIDST toolbox, shown in Figure 4.1, have already been introduced in the previous deliverables. Deliverable 4.1 [4] described the status of the software development in respect to the Learning Engine component including BN structural and parameter learning as well as the parallelization of structural learning using parallel TAN and PC implementations. Deliverable 2.3 [5] provided more details about the data structures related to the PGM component and the different considered data source management functionalities, namely, DataOnMemory, DataOnDisk, DistributedData, DataStream, DataInstance, and DynamicDataInstance.

In this section, we focus on the Inference Engine component for which we will present in what follows the considered derived components, namely, Variational Message Passing, Importance Sampling, Expectation Propagation, and HUGIN Exact Inference.

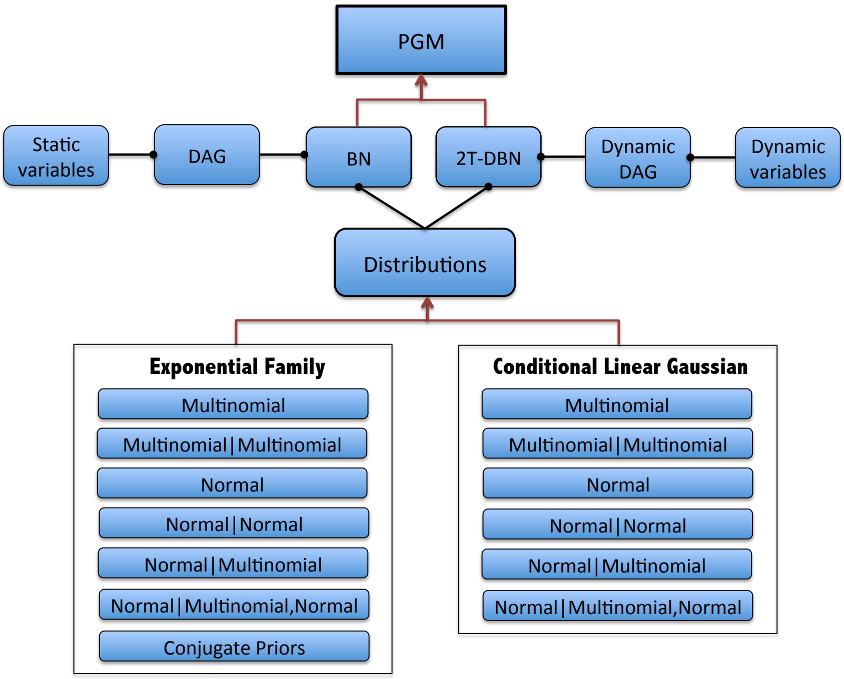

Figure 3.1: Illustration of AMIDST toolbox data structure components. Nomenclature: The boxes in the figure represent software components (sets, possibly singletons, of classes), a rounded-arc going from X to Y indicates that Y 'uses/references' X, and an arc with an arrow from X to Y implies inheritance.

4.1 Variational Message Passing (VMP)

A general architecture for supporting variational message passing (VMP) in graphical models is presented in [6], highlighting how distributions that are conjugate-exponential families [3, 6, 9] can be utilised to efficiently represent the messages by the expected natural statistics. A similar scheme can also be deployed for expectation propagation, but there relying on a transformation between the exponential family representation?s expected sufficient statistics and the distribution?s moments.

4.2 Importance Sampling (IS)

Importance Sampling (IS) [7] is based on the Hybrid Loopy Belief propagation.

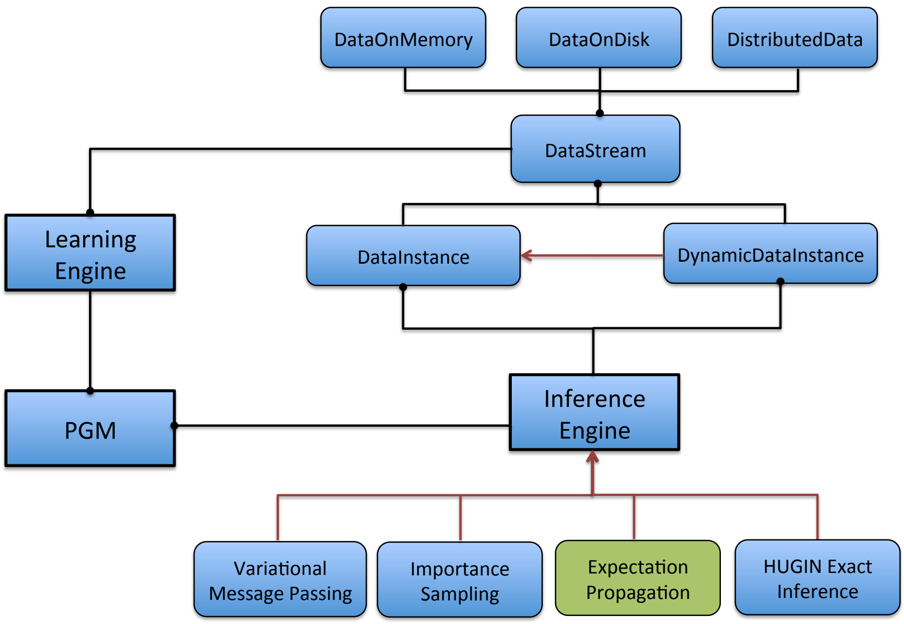

Figure 4.1: Illustration of AMIDST toolbox main core components. Nomenclature: The boxes in the figure represent software components (sets, possibly singletons, of classes), a rounded-arc going from X to Y indicates that Y 'uses/references' X, and an arc with an arrow from X to Y implies inheritance.

4.3 Expectation Propagation (EP)

Expectation propagation (EP) [8] presents a similar approximation scheme as VMP, but it rely more on a transformation between the exponential family representation's expected sufficient statistics and the distribution's moments.

4.4 HUGIN Exact Inference

5 Conclusion

References

[1] Langseth, H., Madsen, A.L., Nielsen, T.D., Rumí, R., Salmerón, A.: State of the art of inference in hybrid and dynamic models (October 2014) Deliverable 3.1 of the

AMIDST project, `amidst.eu`.

[2] Jordan, M., Ghahramani, Z., Jaakkola, T., Lawrence, S.: An introduction to variational methods for graphical models. Machine Learning **37**(2) (1999) 183–233

[3] Attias, H.: A variational Bayesian framework for graphical models. Advances in neural information processing systems (2000) 209—215

[4] Borchani, H., Langseth, H., Nielsen, T.D., Salmerón, A., Madsen, A.L.: Progress report on the software development (December 2014)

[5] Borchani, H., Fernández, A., Langseth, H., Madsen, A.L., Martínez, A.M., Masegosa, A., Nielsen, T.D., Salmerón, A.: A software library implementation of the modellling framework (March 2015) Deliverable 2.3 of the AMIDST project, `amidst.eu`.

[6] Winn, J., Bishop, C.: Variational message passing. Journal of Machine Learning Research **6** (2005) 661–694

[7] Yuan, C., Druzdzel, M.: Importance sampling for general hybrid Bayesian networks. In: Proceedings of the 11th International Conference on Artificial Intelligence and Statistics. (2007) 652–659

[8] Minka, T.: Expectation propagation for approximate Bayesian inference. In: Proceedings of the Seventeenth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-01). (2001) 362–369

[9] Beal, M.J.: Variational algorithms for approximate Bayesian inference. PhD thesis, Gatsby Computational Neuroscience Unit, University College London (2003)