

Bayesian Network Tools in Java (BNJ) v2.0

William H. Hsu

Roby Joehanes

Haipeng Guo

Benjamin B. Perry

Julie A. Thornton

Other Contributors

Prashanth Boddhireddy

Siddharth Chandak

Charles Thornton

<http://bndev.sourceforge.net>

What is BNJ?

- Software toolkit for research and development using graphical models
- Open source (GNU General Public License)
- 100% Java (J2EE v1.4)
- Developed at KDD Lab, Kansas State University
- <http://bndev.sourceforge.net>
- Version 2 currently in alpha stage



Intended Users

- Researchers / students
 - Experiment with algorithms for learning, inference
 - Standardized comparison
 - Synthesis
 - Create, edit, convert **networks, data sets**
- Developers
 - New algorithms for graphical models using BNJ API
 - Applications



Bayesian Network Tools in Java (BNJ) v2.0
<http://bndev.sourceforge.net>



BNJ History

- BNC: initiated 1997, U. Illinois
- BNJ 1: developed 1999-2002, KS State
 - Hard to maintain
 - Redesigned from scratch
- BNJ 2: development started Dec 2002
 - Surpasses BNJ v1 in features, flexibility, performance
 - More standardized API



Bayesian Network Tools in Java (BNJ) v2.0
<http://bndev.sourceforge.net>



BNJ Highlights [1]: Network Interchange

- 8 network formats supported
 - Hugin .net (both 5.7 and 6.0)
 - XML-Bif
 - Legacy BIF
 - Microsoft XBN
 - Legacy DSC
 - Genie DSL
 - Ergo ENT
 - LibB .net
- Opens, saves, converts



BNJ Highlights [2]: Data Formats Supported

- Microsoft Excel (.xls)
- WEKA (.arff)
- LibB data
- XML-data
- Legacy .dat format
- Flat files
 - Space/tab delimited ASCII .txt
 - Comma-separated



BNJ Highlights [3]: Exact Inference

- Junction Tree [Lauritzen & Spiegelhalter, 1988]
- Variable elimination [Shenoy; Dechter] with optimizations
 - *JavaBayes* [Cozman, 2001]
 - Kansas State KDD Lab [Joehanes & Hsu, 2003]
- Singly-connected network belief propagation [Pearl, 1983]
- Cutset Conditioning – under revision [Suermondt, Horvitz, & Cooper, 1990]



Bayesian Network Tools in Java (BNJ) v2.0
<http://bndev.sourceforge.net>



BNJ Highlights [4]: Approximate Inference

- Sampling based:
 - Logic Sampling
 - Forward Sampling
 - Likelihood Weighting
 - Self-Importance Sampling
 - Adaptive Importance Sampling (AIS)
- Bounded Cutset Conditioning (BCC) – under revision
- Hybrid: AIS-BCC bridge – under revision



Bayesian Network Tools in Java (BNJ) v2.0
<http://bndev.sourceforge.net>



BNJ Highlights [5]: Structure Learning

- Greedy (Bayesian Dirichlet) score-based: $K2$ [Cooper & Herskovits, 1992]
- Genetic wrapper
 - cf. [Larranaga, 1998; Hsu, Guo, Perry, Stilson, 2002]
 - GAWK (for $K2$) [Joehanes, 2003]
 - Direct structure learning [Perry, 2003]
- Iterative Improvement
 - Straightforward hill-climbing
 - Simulated annealing (SA)
 - SA with adversarial reweighting
 - Other algorithms



Bayesian Network Tools in Java (BNJ) v2.0
<http://bndev.sourceforge.net>



BNJ Highlights [6]: Analysis and Experimentation

- Structure scoring during, after learning
 - Graph errors
 - RMSE
 - Log likelihood score
 - Dirichlet structure score
- Robustness analysis module
- Data generator: applies existing sampling-based inference algorithms



Bayesian Network Tools in Java (BNJ) v2.0
<http://bndev.sourceforge.net>



BNJ Highlights [7]: Probabilistic Relational Models

- Preliminary support for PRM structure learning
 - Accesses relational databases (*mySQL*, *PostgreSQL*, *ORACLE 9i*) via JDBC interface
 - Preliminary local database loading support (without any database engines)
 - Currently: adapt traditional learning algorithms such as *K2*, *Sparse Candidate*, etc. to relational models
- PRM inference: planned for full release of v2 (Spring, 2004)



BNJ Highlights [8]

- Converter Factory
 - Standalone application
 - GUI front-end
 - Converts among supported network, data formats
- Database GUI Tool
 - Transfer data files to and from server
 - Submit SQL commands through JDBC interface
 - Currently used for PRM learning



BNJ Highlights [9]

- Wizards for
 - Inference*
 - Learning*
 - Others planned
- GUI for Network Editing
 - Still in redevelopment
 - Currently display-mode only
- All tools available in command-line mode



Bayesian Network Tools in Java (BNJ) v2.0
<http://bndev.sourceforge.net>



BNJ Performance

- Relatively fast inference for small to medium networks
- Tends to slow down when node arity high
- Optimization underway
- Very fast learning engine
 - 235 nodes, 76 data points (yeast cell-cycle expression data, Spellman-Gasch) with $K2$: 3 seconds on AMD Athlon XP 1.6GHz
 - Full alarm (37 nodes, 3000 data points) with $K2$: 13 seconds on AMD Athlon XP 1.6GHz



Bayesian Network Tools in Java (BNJ) v2.0
<http://bndev.sourceforge.net>



Applications, New Research: What We Have Done with BNJ

- Computational genomics:
learning gene expression pathways
 - *Saccharomyces cerevisiae* (yeast)
[Johanes & Hsu, 2003]
 - *Oryza sativa* (rice) defense-response – in progress
<http://www.kddresearch.org/REU/Summer-2003>
- PRM Learning Experiments: *EachMovie* data
- New Developments
 - Variable ordering wrappers [Hsu *et al.*, 2002]
 - Hybrid inference algorithms (AIS-BCC)



Bayesian Network Tools in Java (BNJ) v2.0
<http://bndev.sourceforge.net>



Software Demo

- Development using Eclipse platform
 - Open-source IDE
 - From IBM (www.eclipse.org)
- Standalone applications: *coming soon*
- Sources, documentation on *SourceForge*
<http://bndev.sourceforge.net>



Bayesian Network Tools in Java (BNJ) v2.0
<http://bndev.sourceforge.net>



References [1]

■ Applications

- [GHVW98] Grois, E., Hsu, W. H., Voloshin, M., & Wilkins, D. C. (1998). Bayesian Network Models for Automatic Generation of Crisis Management Training Scenarios. In *Proceedings of the Tenth Innovative Applications of Artificial Intelligence Conference (IAAI-98)*, Madison, WI, pp. 1113-1120. Menlo Park, CA: AAAI Press. ([PDF](#) / [PostScript](#) / [.ps.gz](#))

■ General

- [Br95] Brooks, F. P. (1995). *The Mythical-Man Month, 20th Anniversary Edition: Essays on Software Engineering*. Boston, MA: Addison-Wesley.
- [La00] Langley, P. (2000). Crafting papers on machine learning. In *Proceedings of the Seventeenth International Conference on Machine Learning*, Stanford, CA, pp. 1207-1211. San Francisco, CA: Morgan Kaufmann Publishers. ([HTML](#) / [.ps.gz](#))
- [La02] Langley, P. (2002). *Issues in Research Methodology*. Palo Alto, CA: Institute for the Study of Learning and Expertise. Available from URL: <http://www.isle.org/~langley/methodology.html>.



Bayesian Network Tools in Java (BNJ) v2.0

<http://bndev.sourceforge.net>



References [2]

■ Recent and Current Research

- [FGKP99] Friedman, N., Getoor, L., Koller, D., & Pfeffer, A. (1999). Learning Probabilistic Relational Models. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-1999)*, Stockholm, SWEDEN. San Francisco, CA: Morgan Kaufmann Publishers. ([PDF](#))
- [GFTK02] Getoor, L., Friedman, N., Koller, D., & Taskar, B. (2002). Learning Probabilistic Models of Link Structure. *Journal of Machine Learning Research*, 3(2002):679-707. ([PDF](#))
- [GH02] Guo, H. & Hsu, W. H. (2002). A Survey of Algorithms for Real-Time Bayesian Network Inference. In Guo, H., Horvitz, E., Hsu, W. H., and Santos, E., eds. *Working Notes of the Joint Workshop (WS-18) on Real-Time Decision Support and Diagnosis*, AAAI/UAI/KDD-2002. Edmonton, Alberta, CANADA, 29 July 2002. Menlo Park, CA: AAAI Press. ([PDF](#))
- [Gu02] Guo, H. (2002). A Bayesian Metareasoner for Algorithm Selection for Real-time Bayesian Network Inference Problems (Doctoral Consortium Abstract). In *Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI-2002)*, Edmonton, Alberta, CANADA, p. 983. Menlo Park, CA: AAAI Press. ([PDF](#))
- [HGFS02] Hsu, W. H., Guo, H., Perry, B. B., & Stilson, J. A. (2002). A permutation genetic algorithm for variable ordering in learning Bayesian networks from data. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2002)*, New York, NY. San Francisco, CA: Morgan Kaufmann Publishers. ([PDF](#) / [PostScript](#) / [.ps.gz](#)) - **Nominated for Best of GECCO-2002**. [Genetic Algorithms Demo](#) (31 nominees, 160 accepted papers out of 320)



Bayesian Network Tools in Java (BNJ) v2.0

<http://bndev.sourceforge.net>



References [3]

■ Software

- [Mu03] Murphy, K. P. (2003). *Bayes Net Toolbox v5 for MATLAB*. Cambridge, MA: MIT AI Lab. Available from URL: <http://www.ai.mit.edu/~murphyk/Software/BNT/bnt.html>.
- [PS02] Perry, B. P. & Stilson, J. A. (2002). *BN-Tools: A Software Toolkit for Experimentation in BBNs* (Student Abstract). In *Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI-2002)*, Edmondson, Alberta, CANADA, pp. 963-964. Menlo Park, CA: AAAI Press. (PS)

■ Textbooks and Tutorials

- [Mu01] Murphy, K. P. (2001). *A Brief Introduction to Graphical Models and Bayesian Networks*. Berkeley, CA: Department of Computer Science, University of California - Berkeley. Available from URL: <http://www.cs.berkeley.edu/~murphyk/Bayes/bayes.html>.
- [Ne90] Neapolitan, R. E. (1990). *Probabilistic Reasoning in Expert Systems: Theory and Applications*. New York, NY: Wiley-Interscience. (Out of print; [Amazon.com reference](#))
- [Ne03] Neapolitan, R. E. (2003). *Learning Bayesian Networks*. Englewood Cliffs, NJ: Prentice Hall. ([Amazon.com reference](#))



Bayesian Network Tools in Java (BNJ) v2.0

<http://bndev.sourceforge.net>



References [4]

■ Foundational Material and Seminal Research

- [CH92] Cooper, G. F. & Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9(4):309-347.
- [Jo98] Jordan, M. I., ed. (1998). *Learning in Graphical Models*. Cambridge, MA: MIT Press. ([Amazon.com reference](#))
- [LS88] Lauritzen, S., & Spiegelhalter, D. J. (1988). Local Computations with Probabilities on Graphical Structures and Their Application to Expert Systems. *Journal of the Royal Statistical Society Series B* 50:157-224.

■ Theses and Dissertations Related to BNJ

- [Me99] Mengshoel, O. J. (1999). *Efficient Bayesian Network Inference: Genetic Algorithms, Stochastic Local Search and Abstraction*. Ph.D. Dissertation, Department of Computer Science, University of Illinois at Urbana-Champaign, May, 1999. Available from URL: <http://www-kbs.ai.uiuc.edu/web/kbs/publicLibrary/KBSPubs/Thesis/>.



Bayesian Network Tools in Java (BNJ) v2.0

<http://bndev.sourceforge.net>



References [5]

■ Workshops Relevant to BNJ

- [GHHS02] Guo, H., Horvitz, E., Hsu, W. H., and Santos, E., eds. (2002). *Working Notes of the Joint Workshop (WS-18) on Real-Time Decision Support and Diagnosis*, AAAI/UAI/KDD-2002. Edmonton, Alberta, CANADA, 29 July 2002. Menlo Park, CA: AAAI Press. Available from URL: <http://www.kddresearch.org/Workshops/RTDSDS-2002>.
- [HJP03] Hsu, W. H., Joehanes, R., & Page, C. D. (2003). *Working Notes of the Workshop on Learning Graphical Models in Computational Genomics*, International Joint Conference on Artificial Intelligence (IJCAI-2003). Acapulco, MEXICO, 09 Aug 2003. Available from URL: <http://www.kddresearch.org/Workshops/IJCAI-2003-Bioinformatics>.

