Bayesian Network Tools in Java (BNJ) v2.0

William H. Hsu
Roby Joehanes
Haipeng Guo
Benjamin B. Perry
Julie A. Thornton

Other Contributors
Prashanth Bodhireddy
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
- 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

BNJ History

- 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
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]

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
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

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
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

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
Applications, New Research: What We Have Done with BNJ

- Computational genomics: learning gene expression pathways
  - *Saccharomyces cerevisiae* (yeast) [Johanes & Hsu, 2003]

- PRM Learning Experiments: *EachMovie* data

- New Developments
  - Variable ordering wrappers [Hsu et al., 2002]
  - Hybrid inference algorithms (AIS-BCC)

Software Demo

- Development using Eclipse platform
  - Open-source IDE
  - From IBM ([www.eclipse.org](http://www.eclipse.org))

- Standalone applications: coming soon

- Sources, documentation on [SourceForge](http://bndev.sourceforge.net) [http://bndev.sourceforge.net](http://bndev.sourceforge.net)
References [1]

Applications

General

References [2]

Recent and Current Research
References [3]

- **Software**

- **Textbooks and Tutorials**

References [4]

- **Foundational Material and Seminal Research**

- **Theses and Dissertations Related to BNJ**
References [5]

- **Workshops Relevant to BNJ**