Lecture 27

Uncertain Reasoning Discussion (4 of 4); KDD and Data Mining Overview

Monday, March 27, 2000

William H. Hsu
Department of Computing and Information Sciences, KSU
http://www.cis.ksu.edu/~bhsu

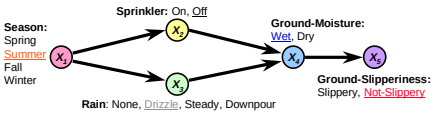
Readings:
"Symbolic Causal Networks for Reasoning about Actions and Plans", Darwiche and Pearl
(Reference) Chapter 15, Russell and Norvig

Lecture Outline

- Readings
  - Chapter 15, Russell and Norvig
  - References
    - Chapters 14-17, Russell and Norvig
    - Chapter 6, Mitchell
    - Pearl and Verma paper
    - Tutorials (Heckerman, Friedman and Goldszmidt)
- Bayesian Belief Networks (BBNs) Concluded
  - Inference: applying CPTs
  - Learning: CPTs from data, elicitation
  - In-class demo: *Hugin* (CPT elicitation, application)
- Causal Discovery and BBN Structure Learning
- KDD and Machine Learning Resources
- Next Class: First KDD Presentation

Bayesian Networks: Quick Review

- Recall: Conditional Independence (CI) Assumptions
- Bayesian Network: Digraph Model
  - Vertices (nodes): denote events (each a random variable)
  - Edges (arcs, links): denote conditional dependencies
- Chain Rule for (Exact) Inference in BBNs $P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))$
  - Arbitrary Bayesian networks: *NP*-complete
  - Polytrees: linear time
- Example ("Sprinkler" BBN)



$P(Summer, Off, Drizzle, Wet, Not\text{-}Slippery) = P(S) \cdot P(O \mid S) \cdot P(D \mid S) \cdot P(W \mid O, D) \cdot P(N \mid W)$

- MAP, ML Estimation over BBNs $h_{ML} = \arg\max_{h \in H} P(D \mid h)$

Learning Structure: State Space Search and Causal Discovery

- Learning Structure: Beyond Trees
  - Problem not as easy for more complex networks
    - Example: allow two parents (even singly-connected case, *aka* polytree)
    - Greedy algorithms no longer guaranteed to find optimal network
    - In fact, *no efficient algorithm exists*
  - Theorem: finding network structure with maximal score, where *H* restricted to BBNs with at most k parents for each variable, is NP-hard for $k > 1$
- Heuristic (Score-Based) Search of Hypothesis Space *H*
  - Define *H*: elements denote possible structures, adjacency relation denotes transformation (e.g., arc addition, deletion, reversal)
  - Traverse this space looking for high-scoring structures
  - Algorithms: greedy hill-climbing, best-first search, simulated annealing
- Causal Discovery: Inferring Existence, Direction of Causal Relationships
  - Want: "No *unexplained correlations*; no *accidental independencies*" (cause $\wedge$ CI)
  - *Can discover causality from observational data alone?*
  - What is *causality* anyway?

Hugin Demo

- *Hugin*
  - Commercial product for BBN inference: http://www.hugin.com
  - First developed at University of Aalborg, Denmark
- Applications
  - Popular research tool for inference and learning
  - Used for real-world decision support applications
    - Safety and risk evaluation: http://www.hugin.com/serene/
    - Diagnosis and control in unmanned subs: http://advocate.e-motive.com
    - Customer support automation: http://www.cs.auc.dk/research/DSS/SACSO/
- Capabilities
  - Lauritzen-Spiegelhalter algorithm for inference (clustering *aka* clique reduction)
  - Object Oriented Bayesian Networks (OOBNs): structured learning and inference
  - Influence diagrams for decision-theoretic inference (utility + probability)
  - See: http://www.hugin.com/doc.html

In-Class Exercise: *Hugin* and CPT Elicitation

- *Hugin* Tutorials
  - Introduction: *causal reasoning for diagnosis in decision support* (toy problem)
    - http://www.hugin.com/hugintro/bbn_pane.html
    - Example domain: explaining low yield (drought versus disease)
  - Tutorial 1: constructing a simple BBN in *Hugin*
    - http://www.hugin.com/hugintro/bbn_tu_pane.html
    - Eliciting CPTs (or collecting from data) and entering them
  - Tutorial 2: constructing a simple influence diagram (decision network) in *Hugin*
    - http://www.hugin.com/hugintro/id_tu_pane.html
    - Eliciting utilities (or collecting from data) and entering them
- Other Important BBN Resources
  - Microsoft Bayesian Networks: http://www.research.microsoft.com/dtas/msbn/
  - XML BN (Interchange Format): http://www.research.microsoft.com/dtas/bnformat/
  - BBN Repository (more data sets) http://www-nt.cs.berkeley.edu/home/nir/public_html/Repository/index.htm

Bayesian Knowledge Discoverer (*BKD*) Demo

- *Bayesian Knowledge Discoverer* (*BKD*)
  - Research product for BBN structure learning: http://kmi.open.ac.uk/projects/bkd/
  - Bayesian Knowledge Discovery Project [Ramoni and Sebastiani, 1997]
    - Knowledge Media Institute (KMI), Open University, United Kingdom
    - Closed source, beta freely available for educational use
  - Handles missing data
  - Uses Branch and Collapse: Dirichlet score-based BOC approximation algorithm http://kmi.open.ac.uk/techreports/papers/kmi-tr-41.ps.gz
- Sister Product: Robust Bayesian Classifier (*RoC*)
  - Research product for BBN-based classification with missing data http://kmi.open.ac.uk/projects/bkd/pages/roc.html
  - Uses Robust Bayesian Estimator, a deterministic approximation algorithm http://kmi.open.ac.uk/techreports/papers/kmi-tr-79.ps.gz

Using ANN, BBN, GA, and ML Tools for KDD

- Learning
  - Bayesian belief networks (BBNs)
    - R. Neal's *DELVE*, MCMC library (University of Toronto)
    - Commercial tools: *Hugin*
    - Experimental: *BKD* (closed-source), *JavaBayes* (open source)
  - Mixture models and Gaussian processes: Neal (Toronto), MacKay (Oxford)
  - Artificial neural network (ANN) tools
    - Commercial (source available): *NeuroSolutions 3*
    - Open source: *Stuttgart Neural Network Simulator (SNNS)*
  - Genetic algorithms (GA) and genetic programming (GP) tools: *Genesis*, *GPSYS*
- Inference
  - BBNs: *Ergo* (MacOS), *Hugin* (Windows)
  - ANNs: NeuroSolutions, SNNS, etc. (see ANN FAQ, *NeuroNet* web page)
- Other KDD Resources
  - *KDNuggets* (http://www.kdnuggets.com)
  - D. Aha's ML page (NRL), AI page (CMU), S. Russell's AIMA page

ANN, BBN, and ML Tools: Questions and Answers

- In-Class Q&A
  - ?

Learning Structure: Conclusions

- Key Issues
  - Finding a criterion for inclusion or exclusion of an edge in the BBN
  - Each edge
    - "Slice" (axis) of a CPT *or* a commitment to acquire one
    - Positive statement of conditional dependency
- Other Techniques
  - Focus today: constructive (score-based) view of BBN structure learning
  - Other score-based algorithms
    - Heuristic search over space of addition, deletion, reversal operations
    - Other criteria (information theoretic, coding theoretic)
  - Constraint-based algorithms: *incorporating knowledge into causal discovery*
- Augmented Techniques
  - Model averaging: optimal Bayesian inference (integrate over structures)
  - Hybrid BBN/DT models: use a decision tree to record $P(x \mid Parents(x))$
- Other Structures: e.g., Belief Propagation with Cycles

Terminology

- Bayesian Networks: Quick Review on Learning, Inference
  - Structure learning: determining the best topology for a graphical model from data
    - Constraint-based methods
    - Score-based methods: statistical or information-theoretic degree of match
    - Both can be global or local, exact or approximate
  - Elicitation of subjective probabilities
- Causal Modeling
  - Causality: "direction" from cause to effect among events (observable or not)
  - Causal discovery: learning causality from observations
- Incomplete Data: Learning and Inference
  - Missing values: to be filled in given partial observations
  - Expectation-Maximization (EM): iterative refinement clustering algorithm
    - Estimation step: use current parameters Θ to estimate missing $\{N_i\}$
    - Maximization (re-estimation) step: update Θ to maximize $P(N_i, E_j \mid D)$

Summary Points

- Bayesian Networks: Quick Review on Learning, Inference
  - Learning, eliciting, applying CPTs
  - In-class exercise: *Hugin* demo; CPT elicitation, application
  - Learning BBN structure: constraint-based versus score-based approaches
  - *K2*, other scores and search algorithms
- Causal Modeling and Discovery: Learning Causality from Observations
- Incomplete Data: Learning and Inference (Expectation-Maximization)
- Tutorials on Bayesian Networks
  - Breese and Koller (AAAI '97, BBN intro): http://robotics.Stanford.EDU/~koller
  - Friedman and Goldszmidt (AAAI '98, Learning BBNs from Data): http://robotics.Stanford.EDU/people/nir/tutorial/
  - Heckerman (various UAI/IJCAI/ICML 1996-1999, Learning BBNs from Data): http://www.research.microsoft.com/~heckerman
- Next Class: BBNs and Causality
- Later: UAI Concluded; KDD, Web Mining; GAs, Optimization