Graphical Models: Bayesian modeling in the large

Dr. Wray Buntine
Heuristicrats Research, Inc.

wray@Heuristicrat.COM
http://www.Heuristicrat.COM/wray/

1678 Shattuck Avenue, Suite 310 • Berkeley, CA, 94709-1631
Tel: +1 (510) 845-5810 • Fax: +1 (510) 845-4405

Section in the tutorial at Maximum Entropy and Bayesian Methods,
Sante Fe, New Mexico, June 31st, 1995.

Outline

• Example graphical models and their analysis
  – a causal and non-causal model
  – causal vs. inference model
  – repeated trials
  – a random walk versus a Markov chain
• Using graphical models
  – applications of Bayesian inference
  – knowledge acquisition and inference in diagnosis
  – compiling a data analysis algorithm from specification
  – a framework for understanding Bayesian inference
• For tutorial references see

Two non-causal models and a causal one
(from Shachter, 199?)

a) bicycle accidents \rightarrow \text{watermelon sales}

b) bicycle accidents \rightarrow \text{watermelon sales}

c) bicycle usage \rightarrow \text{season}

What the graphs mean:

a) \( p(\text{bicycle-acc},\text{waterm-sales}) = p(\text{bicycle-acc})p(\text{waterm-sales}|\text{bicycle-acc}) \)

b) \( p(\text{bicycle-acc},\text{waterm-sales}) = p(\text{waterm-sales})p(\text{bicycle-acc}|\text{waterm-sales}) \)

c) \( p(\text{bicycle-acc},\text{waterm-sales},\text{bicycle-use}, \text{season}, \text{bread-sales}) = p(\text{bread-sales})p(\text{season})p(\text{bicycle-use}|\text{season})p(\text{waterm-sales}|\text{season})p(\text{bicycle-acc}|\text{bicycle-use}) \)

Causality and graphs:

- c) could be said to be “causal models”, but
- a) and b) merely represent probabilities

We can also conclude independence statements:

e.g. from c), \( \text{bicycle-acc independent of waterm-sales given season} \)
\( \text{bicycle-acc not necessarily independent of waterm-sales} \)

Causal vs. Inference models
(See Shachter and Heckerman, 87)

a) \text{occupation} \rightarrow \text{climate} \rightarrow \text{age} \rightarrow \text{disease} \rightarrow \text{symptoms}

b) \text{age} \rightarrow \text{occupation} \rightarrow \text{climate} \rightarrow \text{disease} \rightarrow \text{symptoms}

- a) a causal model
- b) an inference model
- c) a model with NO assumptions

• Inference can work in any direction, i.e., marginalizing and conditioning allows arbitrary probabilities to be inferred.

• Probability tables for the inference model can be derived automatically using Bayes theorem from the causal model (see Shachter, Andersen and Szolovits, 94).
Repeated trials
(see Howard, 70)

- we toss a thumbtack $N$ times with unknown probability $\theta$ of it landing on its flat
- a) gives the evidence (the outcome of each toss), and its probabilistic relationship with $\theta$
  - tosses are independent given $\theta$
  - also called i.i.d. sampling
- b) gives the result of simplifying a) where the tosses are summarized in sufficient statistics $N$ and #flat

Random walk versus a Markov chain

a) today’s price is effected by yesterday’s price only
b) the random walk: daily changes are independent of one another
c) Sincere Trading Co.'s secret hidden Markov model that lets them predict the exchange rate accurately

Outline

- Example graphical models and their analysis
  - a causal and non-causal model
  - causal vs. inference model
  - repeated trials
  - a random walk versus a Markov chain
- Using graphical models
  - applications of Bayesian inference
  - knowledge acquisition and inference in diagnosis
  - compiling a data analysis algorithm from specification
  - a framework for understanding Bayesian inference
- For tutorial references see

Potential applications of Bayesian inference

- million/billion dollar applications in industry waiting for the right software:
  - image understanding, e.g., face or signature recognition
  - document processing and retrieval
  - natural language understanding and intelligent user I-O
  - industrial diagnosis, process control, and instrumentation
  - scientific, medical and sociological data analysis,
    e.g., estimating the number of heroin addicts from indirect sources
  - help desk and intelligent console management
- too many of these to have a CS/NN/Stats/Phys/Eng PhD. work on each one
Knowledge acquisition and inference in diagnosis

- Application areas:
  - Diagnosis of large complex machinery (e.g., turbines), medical diagnosis, computer peripherals (printers), help desk

- Need to acquire knowledge:
  - Key “hidden” variables and “causal” structure
  - 1000's of probabilities (often magnitudes are all that matter)
  - Probabilities are often the expert’s subjective opinions

- Inference is non-directed:
  - Each instance may have different observations
  - May involve hundreds of variables
  - System may be required to recommend which of several (expensive) observations to make next
  - Hypothetical (what-if) reasoning, and explanation

BUGS: compiling a data analysis algorithm

- Bayesian analysis Using Gibbs Sampling.
  - From Gilks, Spiegelhalter, and Thomas (MRC, Cambridge, UK)
  - Takes a data analysis problem represented as a Bayesian Network with Plates and compiles a Gibbs sampler for the problem.

- Amazing variety of problems addressed:
  - Logistic regression
  - Dose response studies
  - Normal mixture models
  - Non-linear regression with heterogeneous variance
  - Discrete variable latent class models
  - Spatial smoothing

- Interfaces to S-Plus, both input and output.

- Gibbs sampling is inherently slow, so useful for smaller samples (e.g., 200 cases), and no multivariate Gaussians in BUGS (!!)

Inferring the distance to galaxies


- Measurements of Cephoids (a class of supergiant variable stars) in a galaxy vary by a constant offset depending on distance.

- To infer distance between galaxies, we can jointly estimate the regression line for the measurements from two galaxies, and the constant offset between them.

- See toy data over page.

- Denote \( dy \) as the constant distance between the regression line for two galaxies.

- Here we:
  - Model the problem as a Bayesian network.
  - Use the BUGS compiler to automatically generate a learning algorithm from the Bayesian network model.
  - Plot the results in S-Plus.
A “Model” for the problem

- Data set is list of records of (x,t,galaxy) where galaxy is a boolean indicating whether the offset dy should be added to the measurement t.
- Model parameters are weights-μ, weights-σ, dy and their priors.

Compiling Autoclass
(by Scott Roy, Heuristicrats Research, Inc., hsr@Heuristicrat.COM)

- Simple 3-variable Autoclass III model in (a)
- Fully parameterised model with priors in (b)
- Roy takes the network in (b), combines it with an EM optimizer and compiles out code as efficient as Autoclass in C.
- System also handles: Bayesian Nearest Neighbor, Bayesian version of LVQ, Autoclass with correlation, Jordan and Jacob’s mixture networks, etc.

Understanding Bayesian inference

- many fields use graphs to represent the structure of a problem, e.g. data flow diagrams for visual programming, neural networks, influence diagrams and decision trees in management science, constraint graphs in OR/CS and analysis methods transfer across these fields
- with graphs, elicitation, analysis and composition/decomposition of the components of the problem are made simpler, i.e., they help the user and the application developer
- probabilistic graphical models provide a natural language for object-oriented specification and construction of software, i.e., they can support prototyping of probabilistic software