Probability and the Web

Ken Baclawski
Northeastern University
VIStology, Inc.
Motivation

- The Semantic Web is a framework for expressing logical statements on the Web.
- It does not specify a standard mechanism for expressing probabilistic statements.
- Use cases can be used to evaluate mechanisms for expressing probability on the Web.
- Use cases drive goals to be achieved by a framework for probability on the Web.
Outline

- Use cases
  - Representative sample
  - Significant overlap among the use cases

- Goals
  - Use case driven
  - Emphasis on interoperability and evaluation
Use Cases

- Communication within a community
- Search within scientific and engineering collections
- Supporting scientific and engineering projects
- Abductive Reasoning
- Information Fusion
- Decision Support
Communication in a community

- Probabilistic statements are fundamental to many communities:
  - Science
  - Engineering
  - Medicine

- Probabilities are meaningful only within the context of a stochastic model, which itself has a context (not necessarily probabilistic).

- Bayesian networks are an example of a stochastic modeling technique for specifying dependencies among random variables.
Search within collections

- Semantic annotation
  - Information retrieval
  - Classification
- Bayesian classifiers
  - Improves classification under uncertainty
  - Must be customized for each search criterion
- Combined technique
  - Medical diagnosis
  - Situation assessment
Project Support

- A large project will produce a large document corpus.
- An engineering or scientific project will produce substantial databases of experimental data.
- Probability is the language for expressing the experimental results.
- There is a need for a common language to integrate the document corpus with the experimental data.
Abductive Reasoning

• Finding the best explanation
• Diagnosis and situation awareness are examples of probabilistic abduction.
• Bayes’ Law is the basis for probabilistic abduction.
• Bayesian networks are a general probabilistic mechanism for probabilistic inference.
  - Causal inference
  - Diagnostic inference
  - Mixed inference
Information Fusion

• Combining information from multiple sources
  - Medicine: meta-analysis
  - Sensor networks: multi-sensor fusion

• Fundamental process for situation awareness
  - Military situation awareness
  - Emergency response management

• State estimation of dynamic systems
  - Kalman filter
  - Dynamic Bayesian network
Ontology Based Fusion Use Case Diagram

Decision Support

- A decision tree can be used for specifying a logical decision.
- Decisions may involve uncertain observations and dependent observations so a simple decision tree will not be accurate.
- Influence diagrams
  - Bayesian network extended with utility functions and with variables representing decisions
  - The objective is to maximize the expected utility.
Goals I

- Shared stochastic models
  - Common interchange format
    - Discrete and continuous random variables
    - Static and dynamic models
  - Ability to refer to common random variables
    - Medical: diseases, symptoms
    - Homeland security: organizations, individuals
  - Context specification

- Stochastic inference
  - Both causal and abductive inference
  - Exact and approximate algorithms
Goals II

- Fusion of models from multiple sources
  - Multi-source fusion
  - Dynamic systems and networks
- Reconciliation and validation
  - Significance tests
  - Sensitivity analysis
  - Uncertainty analysis
  - Consistency checking
- Decision support
Goals III

• Ease of use
  – Bayesian networks
  – Stochastic functions as modules
  – Support for commonly used probability distributions and models
  – Component based construction of stochastic models
  – Design patterns and best practices

• Compatibility with other standards

• Internationalization
Bayesian Networks
Stochastic modeling techniques

- Logic programming
- Data modeling
- Statistics
- Programming languages
- World Wide Web
Logic Programming: ICL

- Independent Choice Logic
  - Expansion of Probabilistic Horn abduction to include a richer logic (including negation as failure), and choices by multiple agents.
  - Extends logic programs, Bayesian networks, influence diagrams, Markov decision processes, and game theory representations.
  - Did not address ease of use
Logic Programming: BLP

- Bayesian Logic Programs
  - Prolog notation for defining BNs
  - No separation of logic and BN.

\[
\begin{align*}
\text{iq}(S) & \mid \text{student}(S). \\
\text{ranking}(S) & \mid \text{student}(S). \\
\text{diff}(C) & \mid \text{course}(C). \\
\text{grade}(S,C) & \mid \text{takes}(S,C). \\
\text{grade}(S,C) & \mid \text{iq}(S), \text{diff}(C), \text{takes}(S,C). \\
\text{ranking}(S) & \mid \text{grade}(S,C), \text{takes}(S,C). \\
\text{student}(john). & \text{student}(pete). \\
\text{course(ai).} & \text{course(db).} \\
\text{takes}(john,ai). & \text{takes}(john,db). \text{ takes}(pete,ai). 
\end{align*}
\]
Logic Programming: LBN

- Logical Bayesian Networks (LBN)
  - Separation of logic and BN.

random(iq(S)) <- student(S).
random(ranking(S)) <- student(S).
random(diff(C)) <- course(C).
random(grade(S,C)) <- takes(S,C).

ranking(S) | grade(S,C) <- takes(S,C).
grade(S,C) | iq(S), diff(C).

student(john). student(pete).
course(ai). course(db).
takes(john,ai). takes(john,db). takes(pete,ai).
Data Modeling: PRM

- Probabilistic Relational Model
  - Language based on relational logic for describing statistical models of structured data.
  - Model complex domains in terms of entities, their properties, and the relations between them.
Data Modeling: DAPER

- Directed Acyclic Probabilistic Entity-Relational
  - An extension of the entity-relationship model database structure.
  - Closely related to PRM and the plate model, but more expressive, including the use of restricted relationships, self relationships, and probabilistic relationships.
DAPER Example

DAPER Diagram

Bayesian Network

Data

PRM Diagram
Statistics: Plate Model

- Developed independently by Buntine and the Bayesian inference Using Gibbs Sampling (BUGS) project.
- Language for compactly representing graphical models in which there are repeated measurements.
- Commonly used in the statistics community.
Object-Oriented Bayesian Network

This methodology introduces several notions to BN development:
- Components which can be used more than once
- Groupings of BN nodes with a formally defined interface
  - Encapsulation
  - Data hiding
  - Inheritance
- Inference algorithms can take advantage of the OOBN structure to improve performance
Programming Languages: BLOG

- Bayesian logic
- A first-order probabilistic modeling language under development at UC Berkeley and MIT.
- Designed for making inferences about real-world objects that underlie observed data
  - Tracking multiple people in a video sequence
  - Identifying repeated mentions of people and organizations in a set of text documents.
- Represents uncertainty about the number of underlying objects and the mapping between objects and observations.
World Wide Web

- XML Belief Network (XBN) format developed by Microsoft's Decision Theory and Adaptive Systems Group.
- Bayesian Web (BW)
  - Layered approach
  - Stochastic functions (e.g. BNs, OOBNs) are formally specified on the logical layer.
  - Stochastic operations are on a separate layer.
- PR-OWL
References

BN

ICL

BLP

LBN
PRM

DAPER

OOBN

BLOG
http://people.csail.mit.edu/milch/blog/index.html

Plate Model

XBN

BW

PR-OWL


