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

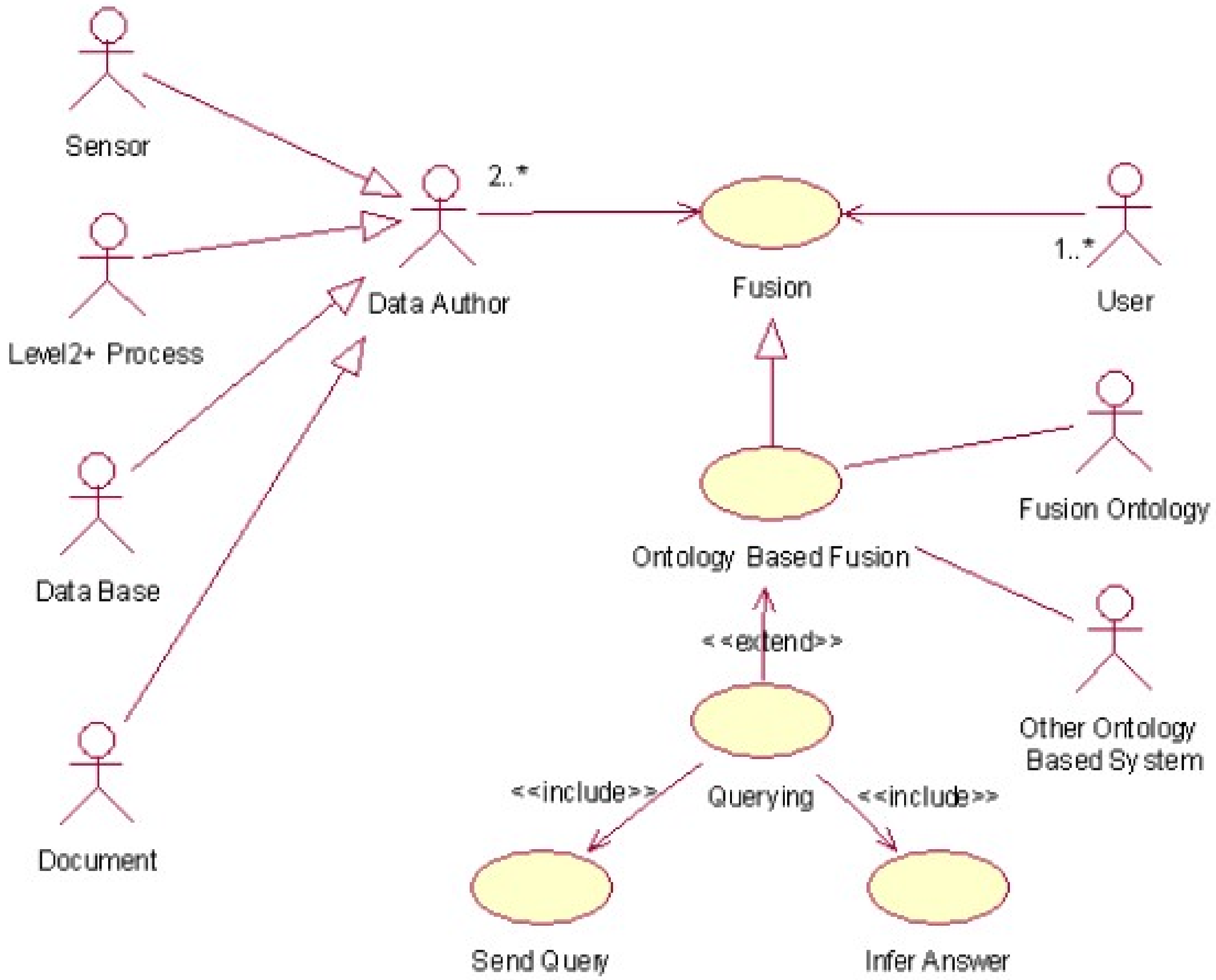
- A large project will produce a large document corpus.
- An engineering or scientific project will produce substantial databases of experimental data.
- Probability is the language for expressing the experimental results.
- There is a need for a common language to integrate the document corpus with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

M. Kokar, C. Matheus, K. Baclawski, J. Letkowski, M. Hinman and J. Salerno. Use Cases for Ontologies in Information Fusion. In *Proc. Seventh Intern. Conf. Info. Fusion*, pages 415-421. (2004)

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

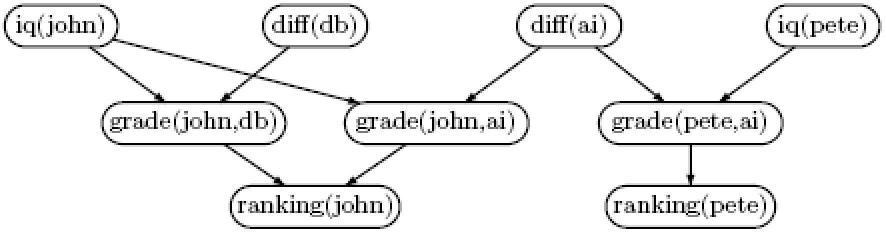
- Bayesian Logic Programs
  - Prolog notation for defining BNs
  - No separation of logic and BN.

```
iq(S) | student(S).
ranking(S) | student(S).
diff(C) | course(C).
grade(S,C) | takes(S,C).
grade(S,C) | iq(S), diff(C), takes(S,C).
ranking(S) | grade(S,C), takes(S,C).

student(john). student(pete).
course(ai). course(db).
takes(john,ai). takes(john,db). takes(pete,ai).
```

Logic Programming: LBN

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



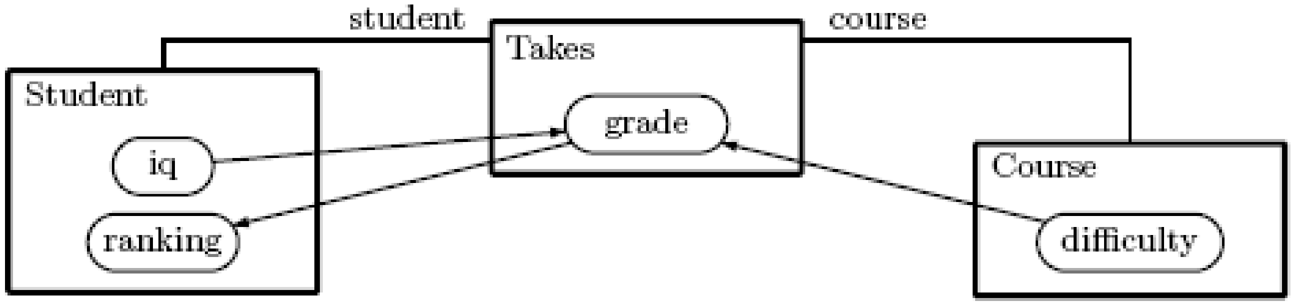

```
random(iq(S)) <- student(S).
random(ranking(S)) <- student(S).
random(diff(C)) <- course(C).
random(grade(S,C)) <- takes(S,C).

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

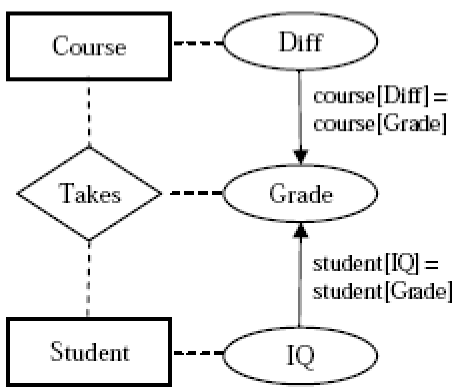
- Probabilistic Relational Model
  - Language based on relational logic for describing statistical models of structured data.
  - Model complex domains in terms of entities, their properties, and the relations between them.



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



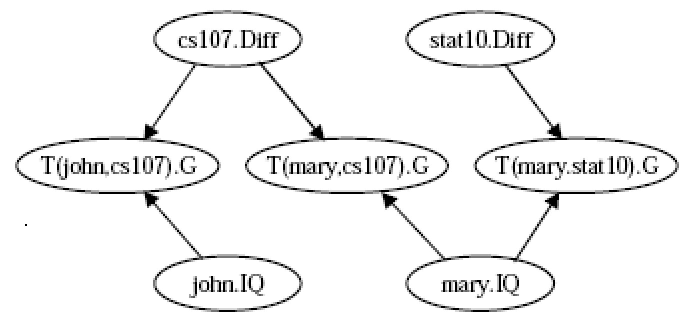
DAPER Diagram

Student
john
mary

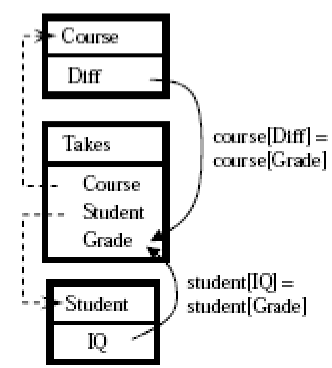
Course
cs107
stat10

Takes	
Student	Course
john	cs107
mary	cs107
mary	stat10

Data



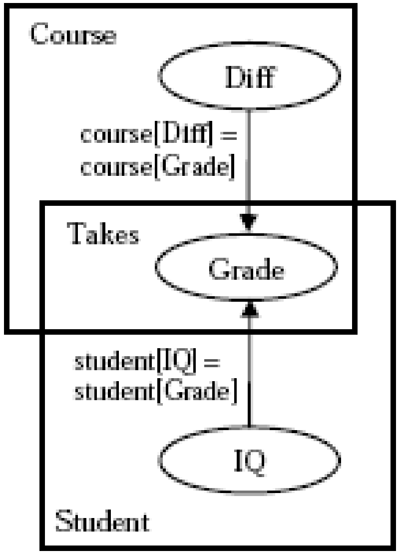
Bayesian Network



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



Programming Languages: OOBN

- 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
Judea Pearl. Fusion, propagation, and structuring in belief networks. Artificial Intelligence 29(3):241-288, 1986.
Judea Pearl. Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann, 1988, ISBN 0-934613-73-7
ICL
D. Poole. Probabilistic Horn abduction and Bayesian networks. Artificial Intelligence, 64:81-129, 1993.
D. Poole. The Independent Choice Logic for modelling multiple agents under uncertainty. Artificial Intelligence, 94(1-2):5-56, 1997.
BLP
K. Kersting and L. De Raedt. Bayesian logic programs. Technical Report 151, Institute for Computer Science, University of Freiburg, Germany, April 2001.
K. Kersting and L. De Raedt. Towards combining inductive logic programming and Bayesian networks. In Proceedings of the 11th International Conference on Inductive Logic Programming (ILP-2001), pages 118-131, 2001.
K. Kersting and U. Dick. Balios - The Engine for Bayesian Logic Programs. In Proceedings of the 8th European Conference on Principles and Practice of Knowledege Discovery in Databases (PKDD-2004), pages 549-551, September 2004.
LBN
H. Blockeel. Prolog for Bayesian networks: a Meta-Interpreter Approach. In Proceedings of the 2nd International Workshop on Multi-Relational Data Mining (MRDM-2003), pages 1-13, 2003.
D. Fierens, H. Blockeel, M. Bruynooghe, and J. Ramon. Logical bayesian networks. In Proceedings of the 3rd Workshop on Multi-Relational Data Mining (MRDM-2004), Seattle, WA, USA, pages 19-30, 2004.
D. Fierens, H. Blockeel, M. Bruynooghe, J. Ramon. Logical Bayesian Networks and Their Relation to Other Probabilistic Logical Models. In S. Kramer and B. Pfahringer (Eds.): ILP 2005, LNAI 3625, pp. 121-135, 2005. Springer-Verlag Berlin, Heidelberg 2005.

PRM
N. Friedman, L. Getoor, D. Koller, and A. Pfeffer. Learning probabilistic relational models. In Proceedings of the 16th International Joint Conference on Artificial Intelligence (IJCAI-1999), pages 1300-1309, 1999.
Learning Probabilistic Relational Models, L. Getoor, N. Friedman, D. Koller, and A. Pfeffer. In Relational Data Mining, S. Dzeroski and N. Lavrac, Eds., Springer-Verlag, 2001
DAPER
D. Heckerman, C. Meek, and D. Koller. Probabilistic Models for Relational Data. Technical Report MSR-TR-2004-30. Microsoft. March 2004.
OOBN
D. Koller, A. Pfeffer. Object-Oriented Bayesian Networks. Proc. 13th Ann. Conf. on Uncertainty in Artificial Intelligence. pp. 302-313. 1997.
BLOG
http://people.csail.mit.edu/milch/blog/index.html
Plate Model
W. Buntine. Operations for learning with graphical models. Journal of Artificial Intelligence Research, 2:159-225. 1994.
C. Spiegelhalter. Bayesian graphical modelling: A case-study in monitoring health outcomes. Applied Statistics, 47:115-134. 1998.
XBN
Microsoft Decision Theory and Adaptive Systems Group. XML Belief Network File Format. http://research.microsoft.com/dtas/bnformat/xbn_dtd.html. April 1999.
BW
K. Baclawski and T. Niu. Ontologies for Bioinformatics. MIT Press. October 2005.
PR-OWL
P. Costa, K. Laskey. PR-OWL: A Framework for Probabilistic Ontologies. Formal Ontologies in Information Systems. 2006.

K. Baclawski, M. Kokar, C. Matheus, J. Letkowski and M. Malczewski. Formalization of Situation Awareness. In *Practical Foundations of Behavioral Semantics*, H. Kilov, K. Baclawski (Ed), pages 25-40. Kluwer Academic. (2003) [pdf]

C. Matheus, K. Baclawski and M. Kokar. Derivation of ontological relations using formal methods in a situation awareness scenario. In *Proc. SPIE Conference on Multisensor, Multisource Information Fusion*, pages 298-309. (April, 2003)

C. Matheus, M. Kokar and K. Baclawski. A Core Ontology for Situation Awareness. In *Proc. Sixth Intern. Conf. on Information Fusion FUSION'03*, pages 545-552. (July, 2003) [pdf]

M. Kokar, C. Matheus, J. Letkowski, K. Baclawski and P. Kogut. Association in Level 2 Fusion. In *Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications*, pages 228-237. (April, 2004) [pdf]

M. Kokar, C. Matheus, K. Baclawski, J. Letkowski, M. Hinman and J. Salerno. Use Cases for Ontologies in Information Fusion. In *Proc. Seventh Intern. Conf. Info. Fusion*, pages 415-421. (2004) [pdf]

C. Matheus, M. Kokar, K. Baclawski, J. Letkowski, C. Call, M. Hinman, J. Salerno and D. Boulware. SAWA: An Assistant for Higher-Level Fusion and Situation Awareness. In *Proc. SPIE Conference on Multisensor, Multisource Information Fusion*, pages 75-85. (2005) [ppt]

C. Matheus, M. Kokar, K. Baclawski, J. Letkowski, C. Call, M. Hinman, J. Solerno and D. Boulware. Lessons Learned from Developing SAWA: A Situation Awareness Assistant. In *Eighth Int. Conf. Info. Fusion* (July 25-29, 2005) [doc]

C. Matheus, K. Baclawski, M. Kokar and J. Letkowski. Using SWRL and OWL to Capture Domain Knowledge for a Situation Awareness Application Applied to a Supply Logistics Scenario. In *Rules and Rule Markup Languages for the Semantic Web First International Conference*, A. Adi, S. Stoutenburg (Ed), pages 130-144. Lecture Notes in Computer Science 3791:130-144. Springer-Verlag. (November 10-12, 2005)

C. Matheus, M. Kokar, K. Baclawski and J. Letkowski. An Application of Semantic Web Technologies to Situation Awareness. In *ISWC'05*, pages 944-958. Lecture Notes in Computer Science 3729:944-958. Springer-Verlag. (2005) [ppt]

M. Kokar, K. Baclawski and H. Gao. Category Theory Based Synthesis of a Higher-Level Fusion Algorithm: An Example. In *Fusion'06* (2006)

M. Kokar, K. Baclawski and C. Matheus. Ontology Based Situation Awareness. *Information Fusion*. to appear. (2006)