Probability and Logic: Bayesian Semantics

Kathryn Blackmond Laskey

George Mason University

Department of Systems Engineering and Operations Research

STIDS 2011 Tutorial Part 3

15 November 2011

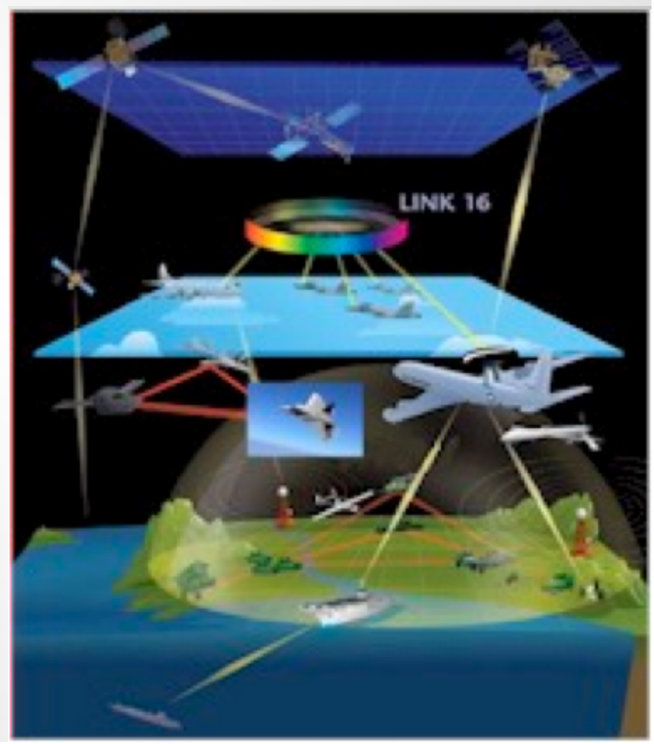
In the olden days…

- We fought big wars
  - Against monolithic enemies
  - Who employed rigid doctrine
  - And fought in predictable ways
- We built stovepipe systems
  - Used by a single organization for a single purpose
  - With idiosyncratic representations and I/O formats
  - Requiring labor-intensive manual transformation of outputs for use by another stovepipe
- Semantics were in the mind of the human
  - Natural language documentation
  - Data structures embedded in code

…and then the world changed.

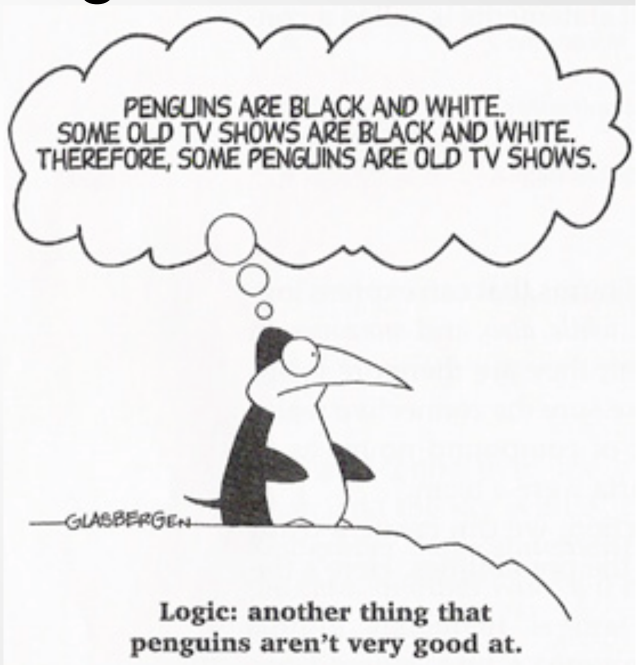
The Age of Semantics

- Today's systems require formal, machine-interpretable semantics
  - Provider and consumer share understanding of inputs and expected outputs
  - Formerly manual functions are fully or partially automated
    - Data interchange
    - Information retrieval
    - Content extraction
    - Discovery of capabilities
  - Users interact with systems at knowledge level



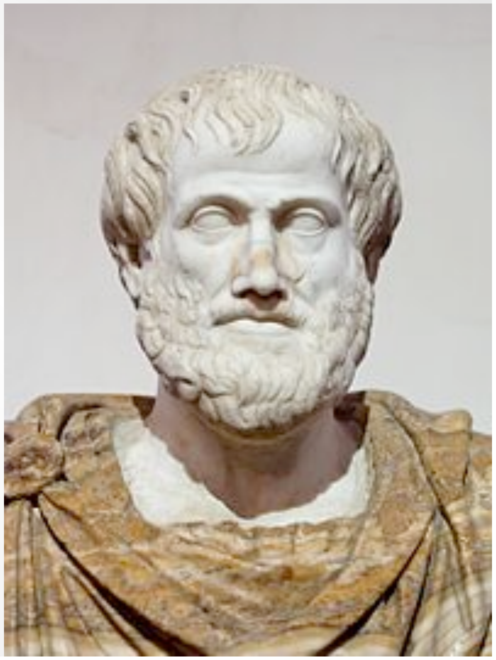
Logic

- Technology for formal, machine-interpretable semantics is founded on logic
- Logic is the study of precise patterns of reasoning
  – Formalize reasoning so it can be carried out automatically
- Russell and Norvig (2002) define a logic as:
  – A formal language for representing knowledge
    - Must have precisely defined syntax and semantics
  – A means of carrying out reasoning in such a language
    - Must have precisely defined reasoning processes that map appropriately to the semantics of the language



Some History

- Aristotle's classical syllogisms (4th century BCE)
- Leibnitz' formalization of Aristotle's syllogisms (17th century)
- Boolean logic (19th century)
- First-order logic by Frege and Pierce (late 19th century)
- Undecidability results, higher order logics, modal logics (20th century)
- Computational logic (late 20th century)



Propositional (Sentential) Logic

- Studies logical relationships among *propositions*
  - A proposition is a declarative statement
  - Complex propositions are built up from elementary propositions using logical connectives
  - Reasoning derives truth-value of conclusions from truth-values of premises
- Insufficiently expressive for expressing semantics of real-world problems
  - Cannot express generalizations
  - Elementary propositions are indivisible units with no inner structure
- But useful as a starting point

p	q	$p \& q$	$p \vee q$	$p \rightarrow q$	$p \leftrightarrow q$
T	T	T	T	T	T
T	F	F	T	F	F
F	T	F	T	T	F
F	F	F	F	T	T

Example: Vehicle Identification

Elementary Propositions:

- » K (Trac<u>k</u>ed vehicle)
- » R (On <u>r</u>oad)
- » F (Traveling <u>f</u>ast)

Axioms:

- » $\neg K \rightarrow R$ (Wheeled vehicle cannot go off-road)
- » $K \rightarrow \neg F$ (Tracked vehicle cannot be traveling fast)

To reason about more than one vehicle, we need to replicate the propositions and axioms "by hand":

$\neg K_i \rightarrow R_i$ *and* $K_i \rightarrow \neg F_i$ *for* i=1, …, N

Possible Worlds

- Axioms define a set of “possible worlds” consistent with axioms
  - Worlds with tracked vehicle traveling fast and wheeled vehicle off-road are impossible
- A truth table uses truth-values of the elementary propositions to determine which worlds are possible

K	R	F	$\neg K \rightarrow R$	$K \rightarrow \neg F$
~~T~~	~~T~~	~~T~~	~~T~~	~~F~~
T	T	F	T	T
~~T~~	~~F~~	~~T~~	~~T~~	~~F~~
T	F	F	T	T
F	T	T	T	T
F	T	F	T	T
~~F~~	~~F~~	~~T~~	~~F~~	~~T~~
~~F~~	~~F~~	~~F~~	~~F~~	~~T~~

Uncertainty is Ubiquitous

- There are many kinds of uncertainty, including:
  – Noise in sensors
  – Intrinsic unpredictability of complex processes
  – Incorrect, incomplete, deceptive intelligence reports
  – Poor understanding of cause and effect relationships
- Representing and reasoning with uncertainty is essential
- Traditional semantic technology provides no support for uncertainty management

"Traditional or deductive logic admits only three attitudes to any proposition: definite proof, disproof, or blank ignorance." - Jeffreys

Probability

- Invented by Laplace, Bernoulli, Bayes in late 18th century
- Many theoretical developments and practical applications in 19th and 20th centuries
  - Formalized set theoretically by Kolmogorov in 1933
  - Formalized as extension to propositional logic by Cox in 1946
- Intense debate on what probability means
- Early work (pre 1985) in knowledge representation ignored probability
- Probabilistic KR became very active after introduction of graphical probability models

Disclaimer

- This talk focuses on *probability*
- There are other ways to generalize logic to obtain truth-values intermediate between proof and disproof
- Other formalisms are also being integrated with semantic technology
- One prominent example is fuzzy logic

Probability	Fuzzy logic
• Attach numerical value to ordinary "crisp" proposition • The proposition is (or will be) definitely true or definitely false • Use traditional logical connectives with usual meaning • Example: "There is a probability of 70% that Mary is between 22 and 28 years old."	• Attach numerical value to "fuzzy" proposition • The proposition has a truth-value intermediate between true and false • Generalize logical connectives to combine degrees of truth • Example: "Mary has a 70% membership in the set of young adults.

GEORGE MASON UNIVERSITY

Interpretations of Probability

1. *Classical* - Ratio of favorable cases to total (equipossible) cases
2. *Frequency* - Limiting value as the number of trials becomes infinite of the frequency of occurrence of some type of event
3. *Logical* - Logical property of one's state of information about a phenomenon
4. *Propensity* - Propensity for certain kinds of event to occur in nature
5. *Subjective* - Ideal rational agent's degree of belief about an uncertain event
6. *Game Theoretic* - Agent's optimal "announced certainty" for an event in a multi-agent game in which agents receive rewards that depend on forecasts and outcomes

1-4 are ontological; 5-6 are epistemological

Possible and Probable Worlds

- Propositional logic can be extended to incorporate uncertainty by assigning a probability to each possible world

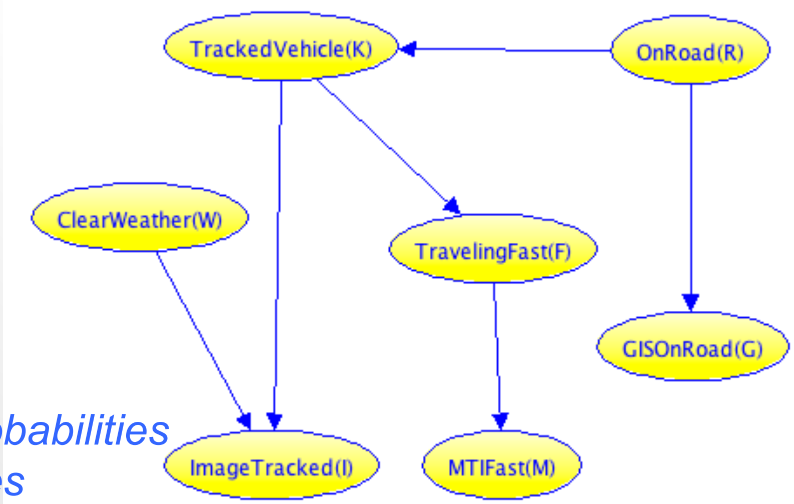
K	R	F	¬K→R	K→¬F	F∨R	Prob
~~T~~	~~T~~	~~T~~	~~T~~	~~F~~	T	0
T	T	F	T	T	T	7.5%
~~T~~	~~F~~	~~T~~	~~T~~	~~F~~	T	0
T	F	F	T	T	F	25%
F	T	T	T	T	T	54%
F	T	F	T	T	T	13.5%
~~F~~	~~F~~	~~T~~	~~F~~	~~T~~	T	0
~~F~~	~~F~~	~~F~~	~~F~~	~~T~~	F	0

Total Probability = 100%

Pr(F∨R) = 7.5% + 54% + 13.2% = 75%

Combinatorics and Graphical Models

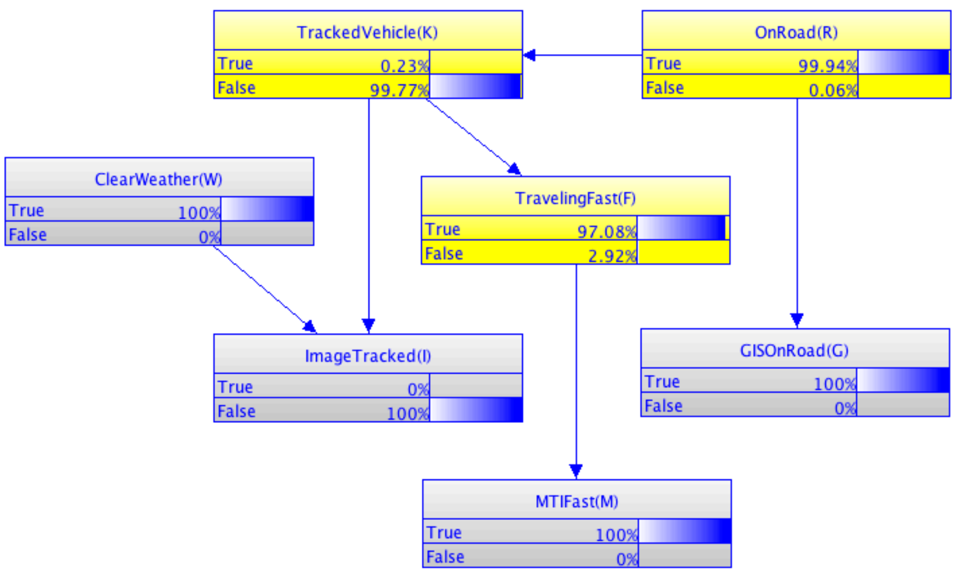
- While theoretically straightforward, combinatorics can be prohibitive
  - With 79 propositions we have $2^{79} = 6.04\text{x}10^{23}$ probabilities (more than Avogadro's number)!
- Enter graphical probability models
  - Graph encodes dependencies among propositions
  - Numerical probabilities specified for a few propositions at a time
  - Tractable specification and inference for problems with thousands of propositions

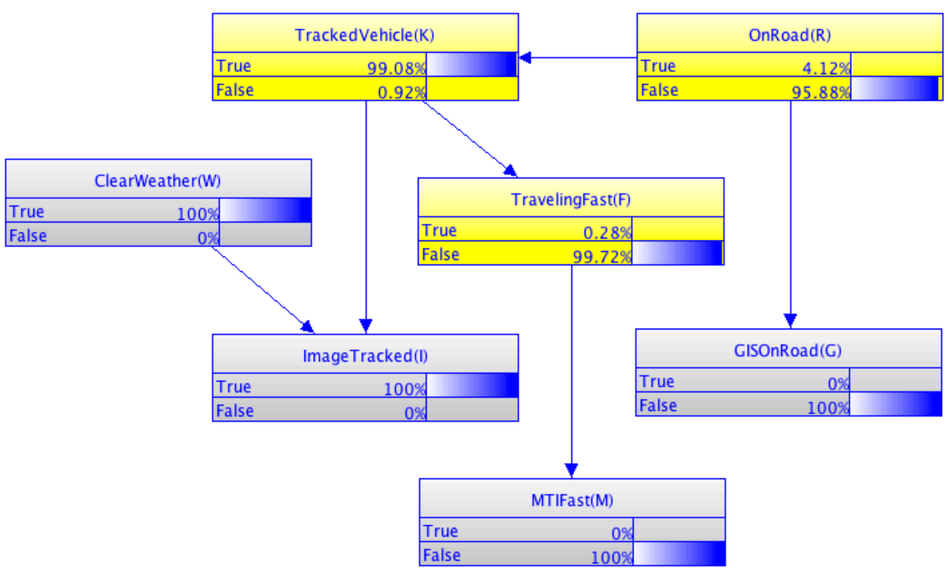


Brute-force specification: 127 probabilities
Bayesian network: 14 probabilities

Probabilistic Reasoning

- Probabilistic reasoning generalizes logical proof
  - Provable propositions have 100% probability
  - Incorporates knowledge falling short of proof
  - Accrues evidence incrementally via Bayes Rule
- Many available off-the-shelf tools





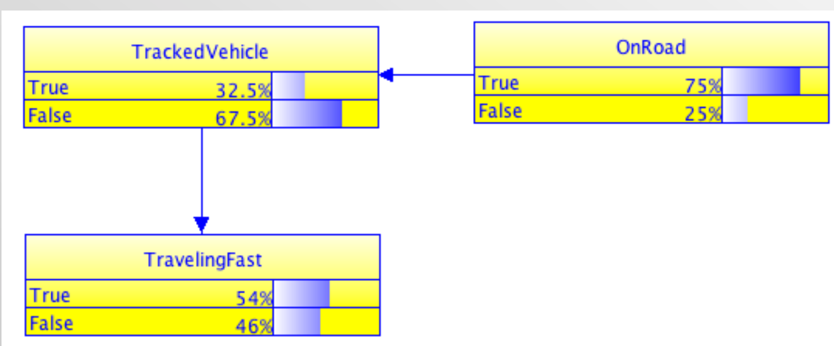
Incorporating Evidence

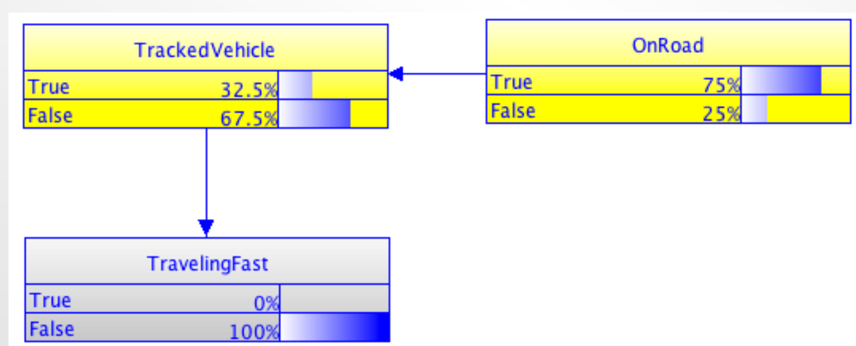
K	R	F	¬K→R	K→¬F	Prob
~~T~~	~~T~~	~~T~~	~~T~~	~~F~~	0
T	T	F	T	T	16.3%
~~T~~	~~F~~	~~T~~	~~T~~	~~F~~	0
T	F	F	T	T	54.35%
~~F~~	~~T~~	~~T~~	~~T~~	~~T~~	0
F	T	F	T	T	29.35%
~~F~~	~~F~~	~~T~~	~~F~~	~~T~~	0
~~F~~	~~F~~	~~F~~	~~F~~	~~T~~	0

Pr(K)=32.5%

Pr(K|¬F) = 70.65%

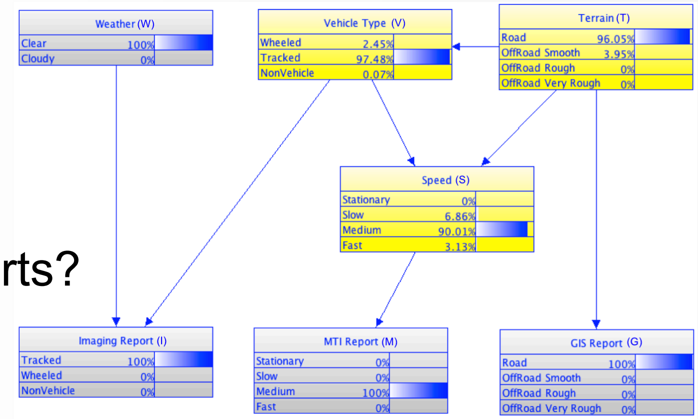
We learn vehicle is not traveling fast

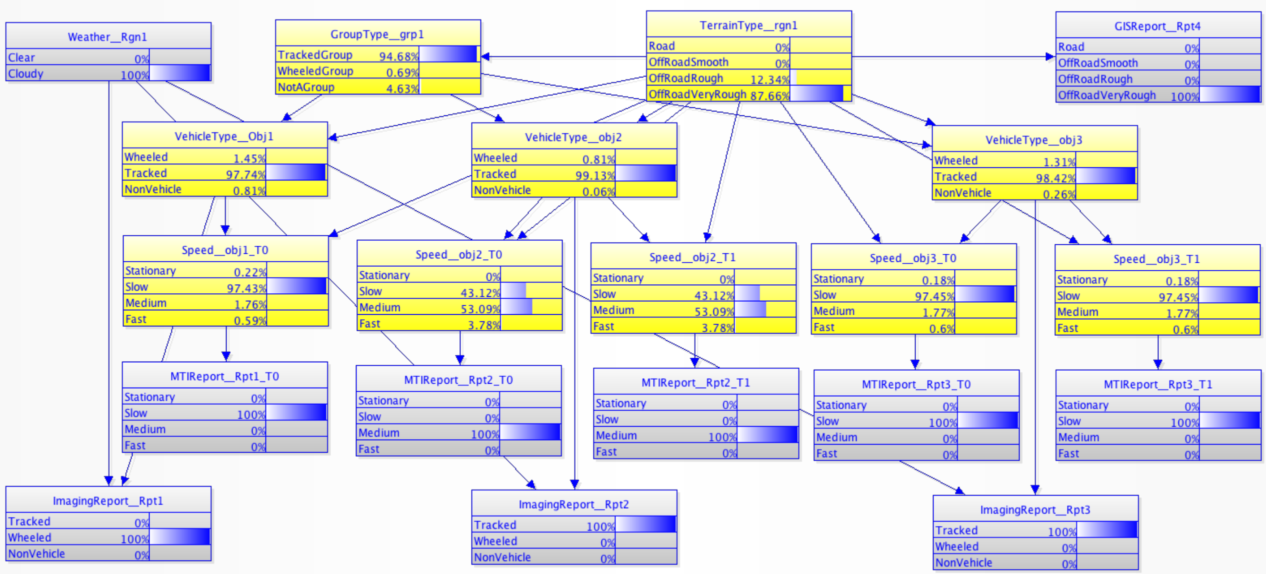




What a BN Cannot Represent

- Repeated structure
  - Different types of entity
  - Multiple entities of each type behave similarly
- Entities are related
  - Which vehicles go with which reports?
  - Are there unreported vehicles? Spurious reports?
- Situation evolves in time
  - Vehicles move
  - New reports arrive





GEORGE MASON UNIVERSITY

First-Order Logic

- Extends expressive power of propositional logic
  - Propositions have inner structure
  - Can express generalizations
    - » For all numbers n and m, n+m is equal to m+n
    - » There is an air defense site next to every airport
    - » No wheeled vehicle can travel off-road
- First-order logic is to propositional logic as algebra is to arithmetic
- Most ontology languages are based on some fragment of first-order logic

Vehicles Revisited: FOL Version

- Propositions have inner structure

  V(x) : x is a vehicle

  L(x) : location of x

  K(x) : x is tracked

  F(x) : x is traveling fast

  R(x) : x is a road

- Can represent:
  - Different types of entity, e.g., vehicles and roads
  - Relationships among entities
  - Functional relationships, e.g., location of object
  - Rules that apply to all entities of a given type, e.g.:
    - $\forall x\ V(x) \rightarrow K(x)\ \neg F(x)$
    - $\forall x\ V(x) \rightarrow \neg K(x) \rightarrow R(L(x))$
  - Particular individual entities, e.g., O_3, O_7
  - Equality, e.g., $O_3=L(O_7)$

First-Order Possible Worlds

- A world for a first-order vocabulary consists of:
  - A set D called the *domain*
    - e.g.. vehicles, roads, and possibly other things
  - An element of D for each constant symbol
    - e.g., a road for O_3 and a vehicle for O_7
  - A relation on D for each relation symbol
    - e.g., the set of objects which are roads for R(x)
  - A function taking arguments in D and having value in D for each function symbol
    - E.g., a function mapping each object to a location (road or non-road) for L(x)
- A world is *possible* for a set of axioms if all the axioms are true in the world

Aside: Terminological Confusion

- A logician calls a possible world for a set of axioms a *model* of the axioms
  - Logician writes axioms representing Newton's laws
  - Logician calls a world in which objects obey Newton's laws a model of the axioms
- An engineer calls a set of equations that are true in a domain a *model* of the domain
  - Engineer writes equations expressing Newton's laws
  - Engineer calls these equations a model of a world in which objects obey Newton's laws
- To avoid confusion, I use "possible world" rather than "model"

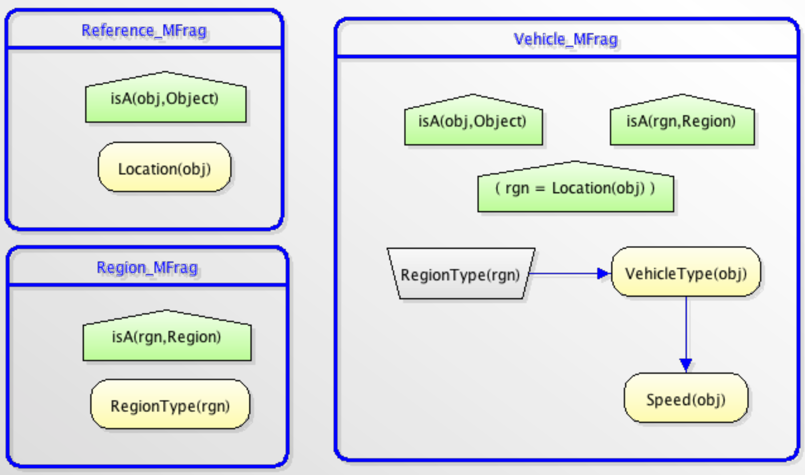
Probability and FOL

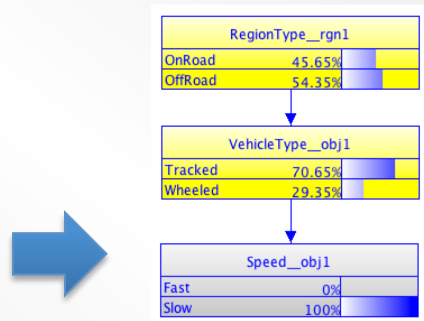
- We extended propositional logic to express knowledge intermediate between proof and disproof
- We would like to do the same with FOL
- We do this by assigning probabilities to sets of possible worlds
- Doing this consistently and tractably is a challenge
- History:
  - Carnap (1950) developed a probability logic for a restricted language
  - Gaifman (1964) developed general theory of probability on sets of first-order possible worlds
  - Research on expressive computational probability logics became active in the late 1990's
- Few usable tools yet exist for the practitioner

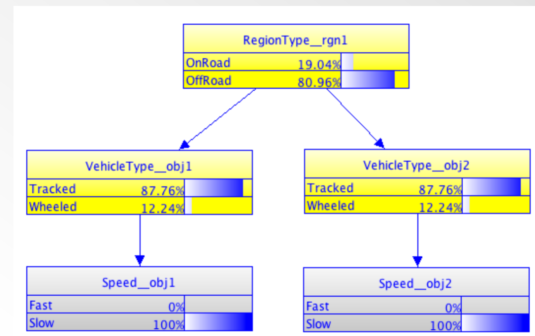
Multi-Entity Bayesian Networks

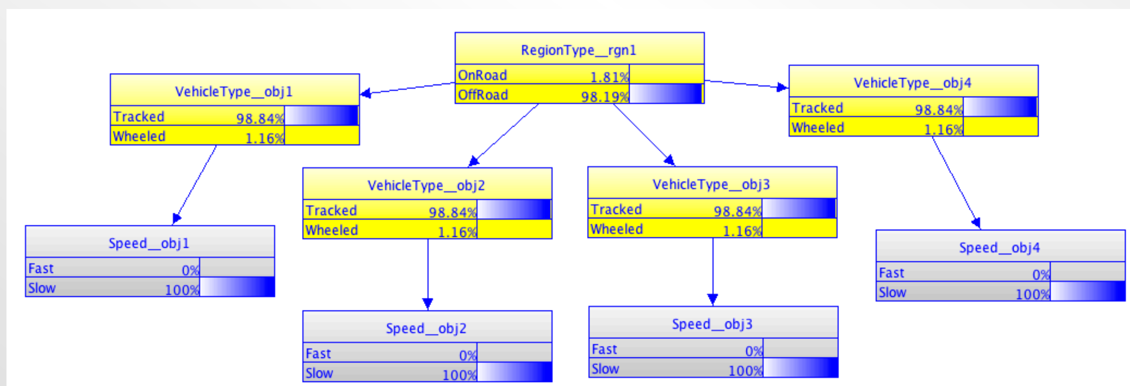
(Laskey, 2008)

- A first-order probabilistic logic
- Represent knowledge as "parameterized BN fragments"
- Instantiate to reason about specific situations









Markov Logic Networks
(Richardson and Domingos, 2006)

Table I. Example of a first-order knowledge base and MLN. Fr() is short for Friends(), Sm() for Smokes(), and Ca() for Cancer().

English	First-Order Logic	Clausal Form	Weight
Friends of friends are friends.	$\forall x \forall y \forall z\ \mathtt{Fr}(x,y) \wedge \mathtt{Fr}(y,z) \Rightarrow \mathtt{Fr}(x,z)$	$\neg\mathtt{Fr}(x,y) \vee \neg\mathtt{Fr}(y,z) \vee \mathtt{Fr}(x,z)$	0.7
Friendless people smoke.	$\forall x\ (\neg(\exists y\ \mathtt{Fr}(x,y)) \Rightarrow \mathtt{Sm}(x))$	$\mathtt{Fr}(x, g(x)) \vee \mathtt{Sm}(x)$	2.3
Smoking causes cancer.	$\forall x\ \mathtt{Sm}(x) \Rightarrow \mathtt{Ca}(x)$	$\neg\mathtt{Sm}(x) \vee \mathtt{Ca}(x)$	1.5
If two people are friends, either	$\forall x \forall y\ \mathtt{Fr}(x,y) \Rightarrow (\mathtt{Sm}(x) \Leftrightarrow \mathtt{Sm}(y))$	$\neg\mathtt{Fr}(x,y) \vee \mathtt{Sm}(x) \vee \neg\mathtt{Sm}(y)$,	1.1
both smoke or neither does.		$\neg\mathtt{Fr}(x,y) \vee \neg\mathtt{Sm}(x) \vee \mathtt{Sm}(y)$	1.1

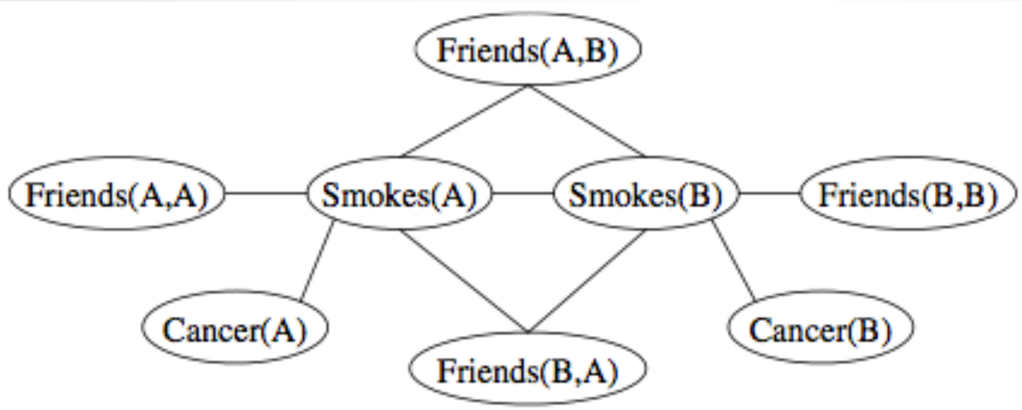


Figure 1. Ground Markov network obtained by applying the last two formulas in Table I to the constants Anna(A) and Bob(B).

Probability Logic

- Many first-order probabilistic languages have recently been developed (c.f., Milch and Russell, 2007)
- Languages draw on different metaphors
  - Database metaphor – probabilistic relational models
  - OO metaphor – object-oriented Bayesian networks
  - Logic metaphor – multi-entity Bayesian networks, Markov logic networks
  - Random variable metaphor – plates
- All these languages can be viewed as defining probabilities on first-order possible worlds

Probability and Ontologies

- A computational ontology (little o) represents
  - Types of objects
  - Relationships among objects
  - Properties of objects
  - Processes and events involving objects
- A *probabilistic ontology** also represents uncertainty about objects, relationships, properties, processes and events
  - Assigns probabilities to possible worlds
  - Respects semantics of the deterministic part of the ontology

**Some object to using the word "ontology" for probabilistic knowledge. Regardless of label, there is a need for integrating probability with semantic technology*

Canonical Reasoning Problems

- **Property value uncertainty** – what is the probability that an individual has a property with a given value?
- **Type uncertainty** – what is the probability that an individual belongs to a class?
- **Reference uncertainty** – which individual plays a given role?
- **Identity uncertainty** – what is the probability that two names refer to the same individual?
- **Existence uncertainty** – what is the probability that a hypothesized individual actually exists?

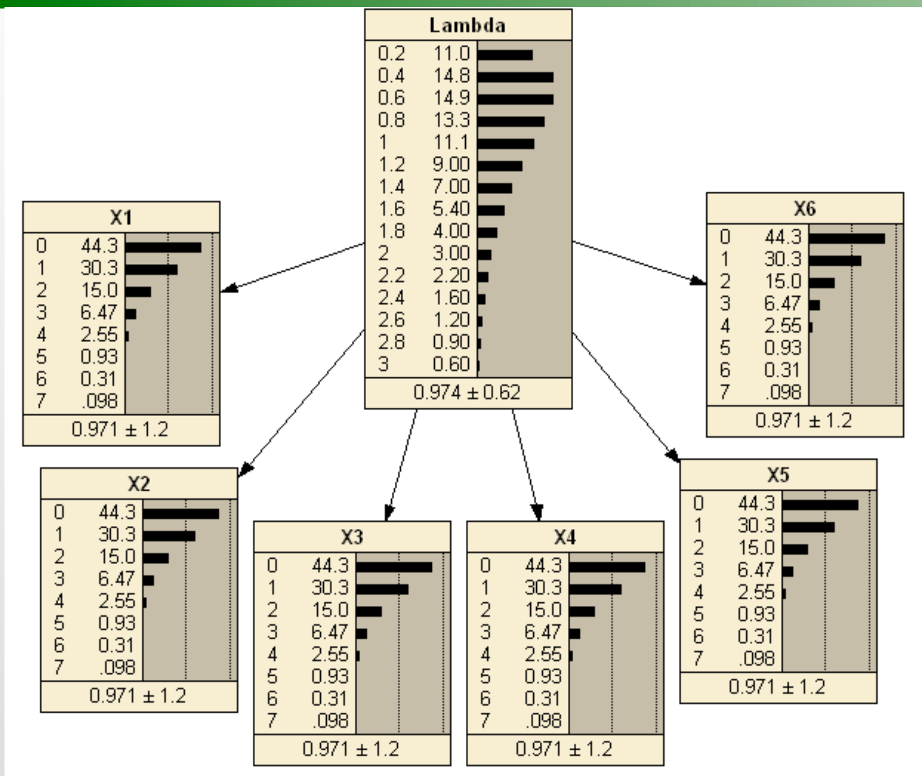
All these can be reduced to property value uncertainty (Poole, et al, 2008)

Learning

- Traditional logic is concerned with *deduction* – deriving logical consequences of a set of axioms
- Probability and statistics are concerned with *induction* – deriving abstractions to generalize observations
- The data mining and machine learning communities are turning to expressive probability logics for powerful, theory-based inductive learning methods
- This class of techniques is known as *Statistical Relational Learning (SRL)*

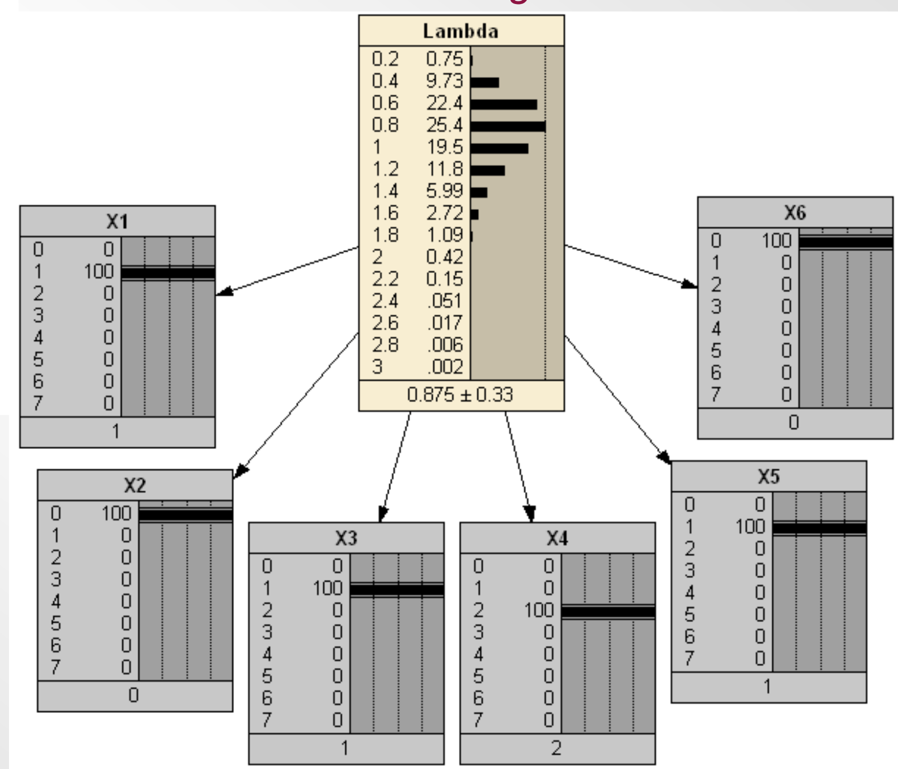
Representing Parameter Learning

Task: Use data from 6 1-hour observation periods to infer rate of transmission errors per hour



Prior distribution for Λ and X

Posterior distribution for Λ given X



Diagrams produced using Netica™ software – available from http:// www.norsys.com

Structure and Parameter Learning

- A probabilistic theory is usually specified by defining:
  - Structure – typically a graph representing dependencies
  - Parameters – typically functions defined on small clusters of propositions representing strength of dependency
- We can represent both structure and parameters explicitly in our knowledge representation and expose them to reasoning
- Thus learning is integral to probability logic

Tractability

- Worst case tractability of FOL + probability is, of course, undecidable
- Efficient exact algorithms exist for restricted (but useful) classes of problems
- Efficient approximations exist for larger classes of problems
- Research is ongoing on characterizing classes of problems and the complexity of methods for those classes

Conclusion

- A theory in classical logic defines a set of possible worlds consistent with the theory
- A probabilistic logic assigns probabilities consistently to sets of possible worlds
- Expressive probability logics
  - Assign probabilities in a way that respects the domain semantics
  - Make use of knowledge that falls short of proof
  - Support evidential accrual
  - Provide built-in learning theory
- Technology for semantically aware uncertainty management is a powerful innovation

References

- R. Carnap, *Logical Foundations of Probability*, University of Chicago Press, Chicago, 1950.
- H. Gaifman, *Concerning measures in First-Order calculi*. Israel Journal of Mathematics, 2, 1–18, 1964.
- Laskey, K.B., MEBN: A Language for First-Order Bayesian Knowledge Bases, *Artificial Intelligence*, 172(2-3): 140-178, 2008
- B. Milch, S. Russell, First-Order probabilistic languages: Into the unknown, in: *Inductive Logic Programming*, Vol. 4455 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 2007, pp. 10–24.
- D. Poole, C. Smyth, R. Sharma, Semantic science: Ontologies, data and probabilistic theories, in: *Uncertainty Reasoning for the Semantic Web I*, Vol. 5327 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 2008, pp. 26–40.
- Richardson, M. and Domingos, P., Markov Logic Networks. *Machine Learning, 62,* 107-136, 2006.

GEORGE MASON UNIVERSITY