Bayesian Knowledge-driven Ontologies

Intuitive Uncertainty Reasoning
for Semantic Networks

Jacob Jurmain
Thayer School of Engineering
Dartmouth College
jacob.c.jurmain@dartmouth.edu
http://di2ag.thayer.dartmouth.edu/

This is joint work with Dr. Eugene Santos, Jr.

Ontologies model a “state of the world”.

- Model a domain in terms of individuals, concepts that describe them, and relationships between them.
  - Description logic (DL) – formal language
- Interesting reasoning task: applying general (terminological) knowledge to case-specific (assertional) knowledge to infer new information.
  - Rule of universal instantiation: “If X is true for each member of a class, then X is true for any particular member of that class.”
  - Build reasoning chains: A → B → C → …

Ontologies' major limitation: they cannot represent uncertainty.

- Uncertainty is multiple possible states of the world, with insufficient knowledge to choose between them.
  - Ex: We have unconfirmed reports about a certain apple that is critical to our national security. It may be red. Or possibly green. If it's green, it's probably a granny smith. But it might just be underripe.
  - How do we describe this with an ontology? We could build one ontology for each possible state of the world, but that is time consuming to build and awful to maintain.
- Uncertainty is everywhere in the real world. Models that can't work with it simply can't address some domains.

Another complication

- Listing the possible states isn't useful. Need some idea of their likelihoods.
  - Ex: There may be some tiny chance of a zombie apocalypse, but probably not enough to warrant civic readiness exercises.

How do we model uncertainty? It's all about variables.

- Variables keep track of the differences between possible world-states.
  - We can describe the possible states of a domain as the cross-product of each variable's possible values.
  - A complete assignment of each variable to one of its values describes a state of the world.

- Knowledge about uncertainty usually concerns how variables interact and influence each other.
  - Ex: $P(A = a_1 | B = b_1) = p$
  - Uncertainty theories use this to compute some measure of likelihood.
    - Probability theory: frequency of event, or degree of belief
    - Fuzzy set theory: partial set membership
    - Possibility theory: possibility and necessity of event

Ontology/uncertainty theories have been attempted, but all have drawbacks.

- Probabilistic Description Logic's (Lukasiewicz, 2008) reasoning is unstable.
  - Inferred rules are progressively less precise than their parents.
- Bayesian Network-Ontology combinations contain a granularity mismatch– BNs' completeness requirement forces them to express knowledge more coarsely than ontologies can.
  - PR-OWL (Costa and Laskey, 2005), BayesOWL (Ding et al, 2005), and P-CLASSIC (Koller et al, 1997) are representative works.
- Fuzzy and possibilistic semantic networks (Straccia, 2001; and Qi et al, 2007; respectively) treat variable interactions coarsely. Information is lost during reasoning.

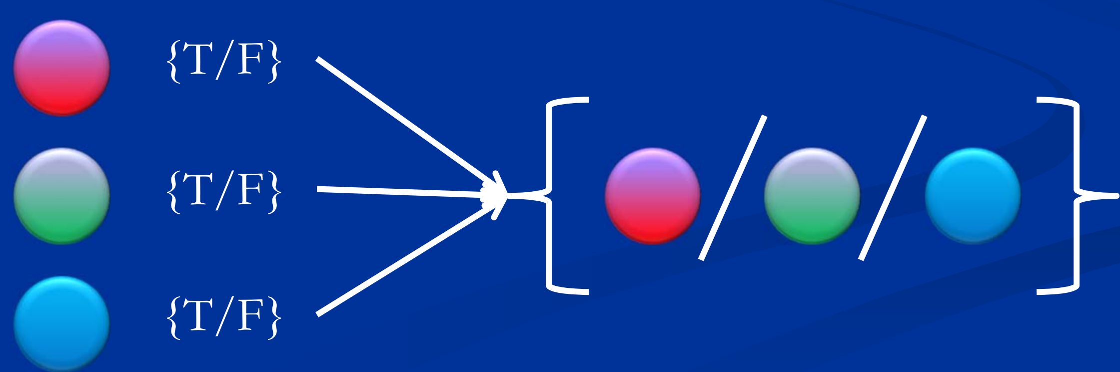
BKOs intuitively overcome these problems.

- BKOs subsume DL ontologies in expression and reasoning.
  - Knowledge is expressed with "if-then" conditional probability rules between assertions.
  - Reasoning is a probabilistic extension of conventional DL reasoning.

- Sophisticated uncertainty reasoning can be done with existing, well-understood tools.
  - Bayesian analyses such as belief updating and revision.
  - Contribution analysis.
  - Fusion of multiple, potentially-conflicting sources.

Three technical insights enable our approach.

1. Consider: for any individual a and any concept C, either $a \in C$ or $a \in \neg C$. Two exclusive states… sounds like a variable.
   - Can define a joint probability distribution over these variables. Equivalently, that distribution is over the ontology's possible states.
   - Can define "constructed variables" from constructed classes. For example, $a \in C \cap D$ or $a \in \neg(C \cap D)$.
   - Ex: A ball that can be any one of three disjoint colors.



Three technical insights (continued)

2. Generalizing the rule of universal instantiation to its probabilistic analog lets DL reason with uncertainty.
   - "If P(X) = p for each member of a class, then P(X) = p for any particular member of that class."

3. BKOs' formulation guarantees conformance to the semantics of a powerful probabilistic modeling/analysis framework called a "Bayesian Knowledge Base" (BKB).

Asserting Knowledge

- Probabilistic equivalents of assertional (case-specific) and terminological (general) axioms.
  - Probabilistic assertional axiom: the probability of one individual's membership in a concept is p, given some other concept memberships [or not – rules can be unconditional].

$$P\left(V_{i_n} = \{a_{i_n} \in B_{i_n}\} \mid \left[V_{i_1} = \{a_{i_1} \in B_{i_1}\} \wedge \ldots \wedge V_{i_{n-1}} = \{a_{i_{n-1}} \in B_{i_{n-1}}\}\right]\right) = p$$

Short form:

$$P\left(a_{i_n} \in B_{i_n} \mid \left[a_{i_1} \in B_{i_1} \wedge \ldots \wedge a_{i_{n-1}} \in B_{i_{n-1}}\right]\right) = p$$

  - Probabilistic terminological axiom: any member of one concept has some probability p of also being a member of another concept.

$$P(x \in D \mid \text{any } x \in C) = p$$

    - C and D can be concept constructor expressions.

BKO example: Mystery Sea Critter, Part 1

- We have grainy, uncertain footage of a newly-discovered ocean critter and are interested in its behavior and biology.

  - Assertional Knowledge: The critter was seen eating something that was either seaweed or a camoflagued seahorse.

$$P(\text{Critter } \textit{ate} \textbf{ some } \text{Seahorse}) = 0.3$$
$$P(\text{Critter } \textit{ate} \textbf{ some } \text{Seaweed}) = 0.6$$

  - Terminological Knowledge: First we classify the critter by diet. Then we consider what that might imply for its eyesight.

$$\text{Seahorse} \sqsubseteq \text{Animal}$$
$$\text{Seaweed} \sqsubseteq \text{Plant}$$

$$P(x \in \text{Carnivore} | \text{any } x \textit{ ate} \textbf{ some } \text{Animal}) = 0.6$$
$$P(x \in \text{Omnivore} | \text{any } x \textit{ ate} \textbf{ some } \text{Animal}) = 0.3$$

$$P(x \in \text{Omnivore} | \text{any } x \textit{ ate} \textbf{ some } \text{Plant}) = 0.5$$
$$P(x \in \text{Herbivore} | \text{any } x \textit{ ate} \textbf{ some } \text{Plant}) = 0.5$$

$$P(x \textit{ eyesight\_quality} \text{ acute} | \text{any } x \in \text{Carnivore} \sqcup \text{Omnivore}) = 0.7$$

Mystery Sea Critter, Part 2

- Original assertional knowledge

  $P(\text{Critter } \textit{ate} \textbf{ some } \text{Seahorse}) = 0.3$ $P(\text{Critter } \textit{ate} \textbf{ some } \text{Seaweed}) = 0.6$

- Instantiating PTAs on "Critter" to generate new inferences:

  $\text{Seahorse} \sqsubseteq \text{Animal}$ ⟶ $P(\text{Critter } \textit{ate} \textbf{ some } \text{Animal} | \text{Critter } \textit{ate} \textbf{ some } \text{Seahorse}) = 1$

  $\text{Seaweed} \sqsubseteq \text{Plant}$ ⟶ $P(\text{Critter } \textit{ate} \textbf{ some } \text{Plant} | \text{Critter } \textit{ate} \textbf{ some } \text{Seaweed}) = 1$

  $P(x \in \text{Carnivore} | \text{any } x \textit{ ate} \textbf{ some } \text{Animal}) = 0.6$
  ⟶ $P(\text{Critter} \in \text{Carnivore} | \text{Critter } \textit{ate} \textbf{ some } \text{Animal}) = 0.6$

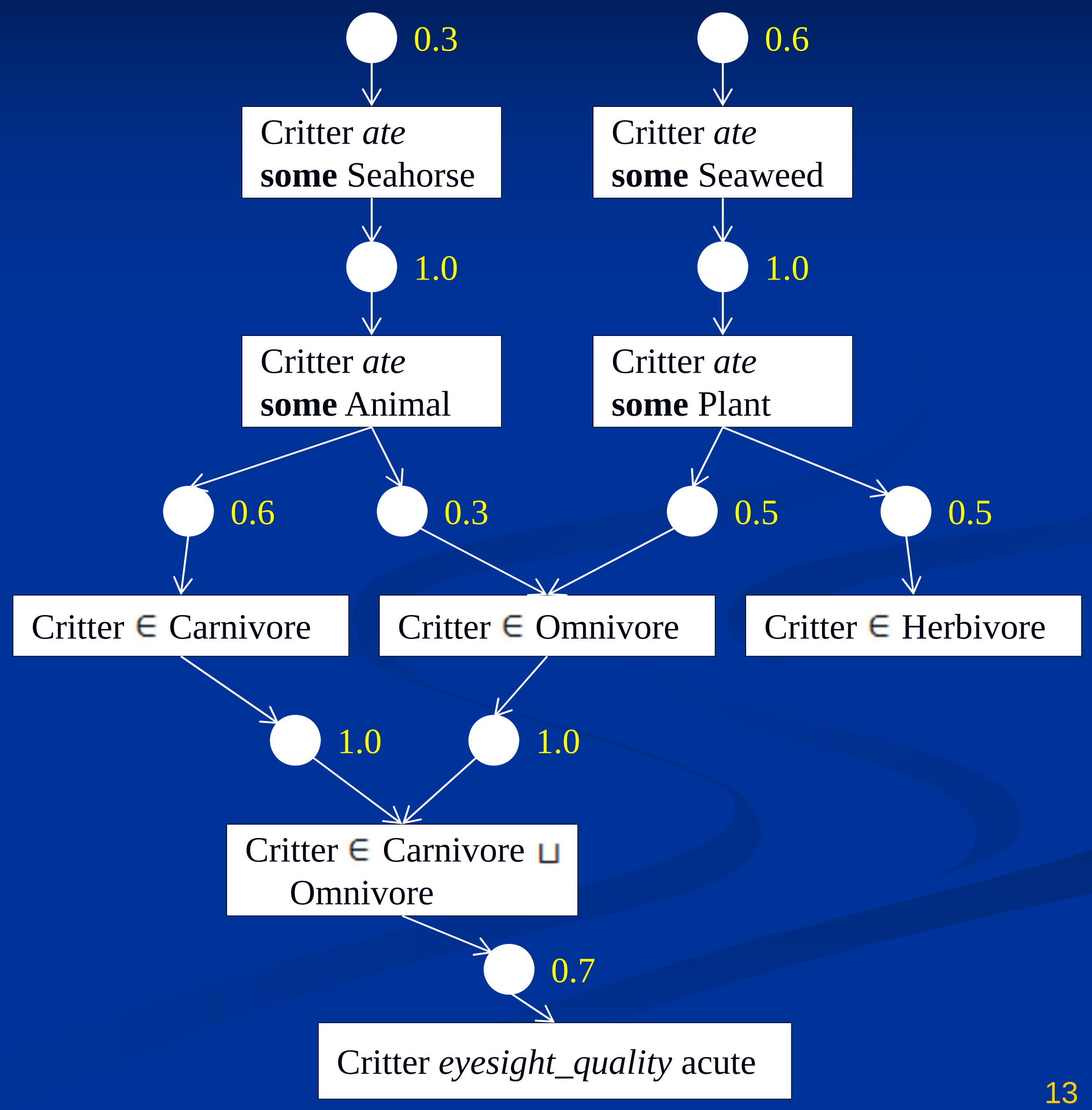
  $P(x \in \text{Omnivore} | \text{any } x \textit{ ate} \textbf{ some } \text{Animal}) = 0.3$
  ⟶ $P(\text{Critter} \in \text{Omnivore} | \text{Critter } \textit{ate} \textbf{ some } \text{Animal}) = 0.3$

  $P(x \in \text{Omnivore} | \text{any } x \textit{ ate} \textbf{ some } \text{Plant}) = 0.5$
  ⟶ $P(\text{Critter} \in \text{Omnivore} | \text{Critter } \textit{ate} \textbf{ some } \text{Plant}) = 0.5$

  $P(x \in \text{Herbivore} | \text{any } x \textit{ ate} \textbf{ some } \text{Plant}) = 0.5$
  ⟶ $P(\text{Critter} \in \text{Herbivore} | \text{Critter } \textit{ate} \textbf{ some } \text{Plant}) = 0.5$

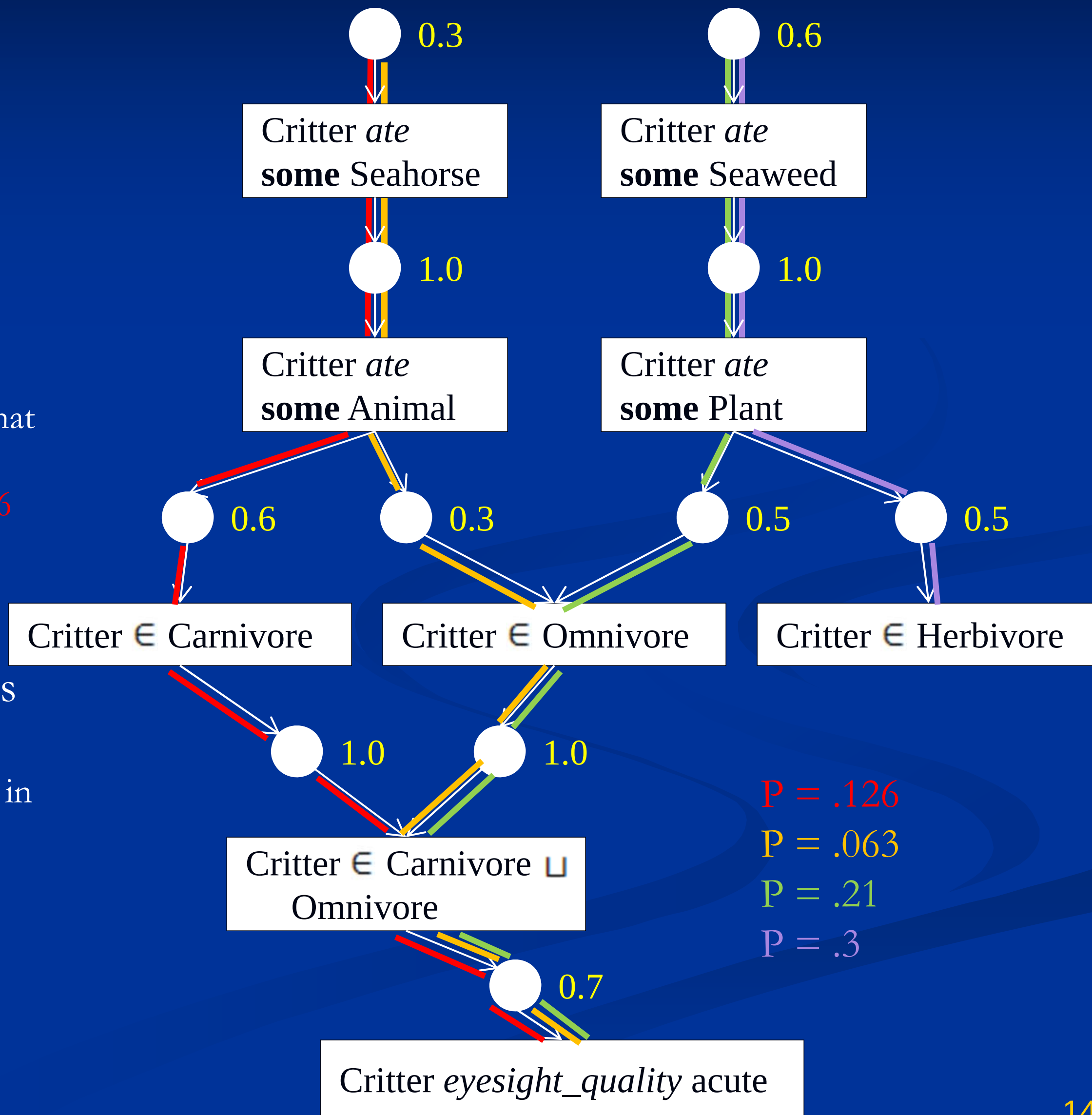
  $P(x \textit{ eyesight\_quality} \text{ acute} | \text{any } x \in \text{Carnivore} \sqcup \text{Omnivore}) = 0.7$
  ⟶ $P(\text{Critter } \textit{eyesight\_quality} \text{ acute} | \text{Critter} \in \text{Carnivore} \sqcup \text{Omnivore}) = 0.7$

Mystery Sea Critter, Part 3



BKB Analyses

- Belief revision: determine most probable state of the world.
  - P = .3
- Belief updating: compute posterior probability of a single variable assignment.
  - Sum of probabilities of inferences that assignment appears in.
  - P(Critter *eyesight_quality* acute) = .126 + .063 +.21 = .399
- Contribution analysis: compute how much one assignment seems to cause another.
  - Sum of probabilities of world states in which the cause appears with the effect, divided by the effect's probability from updating.
  - Contribution of "Critter *ate* **some** Seahorse" to "Critter *eyesight_quality* acute": (.126 + .063) / .399 = .474



Next Development Steps

- Additional theory development to incorporate advanced ontology and BKB capabilities.
  - Ontology expressions beyond DL
  - BKB fusion, sensitivity analysis, interfacing with social networks…
- Optimization of reasoning algorithm.
- Software development.
  - Representation language/format for domain libraries and case models
  - Reasoning application

Additional material

Probabilistic Description Logic

- Founded on Probabilistic Logic (Nilsson, 1986)
- *Expressive Probabilistic Description Logics* (Lukasiewicz, 2008) is representative of the field.

- Assigns probability intervals to DL assertions.
  - Ex: 0.7 ≤ P(Tweety *is-a* **Bird**) ≤ 0.8
  - Not very intuitive. Uncertainty on an uncertainty metric?
  - Limitation: inferred probability intervals' relative precision (width ÷ mean) decays during forward chaining. This cripples deep reasoning.

Ex: 0.7 ≤ P(Tweety *is-a* **Bird**) ≤ 0.8	RP: 0.1/0.75 = 0.13
0.9 ≤ P(**Birds** *can* fly) ≤ 0.99	RP: 0.09/0.945=0.095
→ 0.7 * 0.9 = 0.63 ≤ P(Tweety *can* fly) ≤ 0.8 * 0.99 = 0.79	RP: 0.16/0.71 = 0.22

Bayesian Networks and Ontologies

- Founded on Bayesian Networks (Pearl, 1985)
  - Restricted subclass of Bayesian Knowledge Bases that assumes complete information.
  - BNs require complete definition of "conditional probability tables" instead of working with individual rules like BKBs.

- Defines conditional probability tables using DL assertions as variables.
  - DL does not have BNs' completeness requirement. Using BNs restricts the system's expressiveness.
  - There are notions we can represent in DL that don't work in BNs even when completely known.
  - Ex: Model probability distributions of gas mileage for various airplane models. What happens when one is a glider? Then any distribution, even context-specific independence (Boutilier et al, 1996), is unintuitive.

Fuzzy Description Logic

- Founded on fuzzy logic / fuzzy set theory (Zadeh, 1965)
- *Reasoning within fuzzy description logics* (Straccia, 2001) is a representative work.

- Extends DL to allow partial membership in concepts.
  - Coarse treatment of uncertainty with some information loss during reasoning. Does not intuitively capture if-then interactions like probability theory.
  - Ex: given the assertions

    a in C : 0.7   a in D : 0.4   C in E : 0.2   D in E : 0.6

    what is the membership of a in E?

    max(min(0.7, 0.2), min(0.4,0.6)) = 0.4

    Most of the numbers in the reasoning chain had no effect on the outcome. We usually don't think of causality as working this way.

Possibilistic Description Logic

- Founded on possibility theory (Zadeh, 1978) which extends fuzzy logic.
- *A possibilistic extension for description logics* (Qi et al, 2007) is a representative work.

- Models a DL assertion's uncertainty as two fuzzy numbers, possibility and necessity.
  - Possibility: to what degree could the assertion be true? Necessity: to what degree must the assertion be true?
  - Mathematically, possibility and necessity are simply two fuzzy description logic problems in parallel, with the axiom that possibility $\geq$ necessity.

  - As with fuzzy logic , this is a coarse treatment of causality.