Probabilistic Representation and Reasoning

Alessandro Panella

Department of Computer Science
University of Illinois at Chicago

May 4, 2010

Sources

- R.J. Brachman and H.J. Levesque, Knowledge Representation and Reasoning, Elsevier, 2003
- S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 2nd ed., Prentice Hall, 2003

Outline

1. Introduction
   - Motivations
   - General Concepts

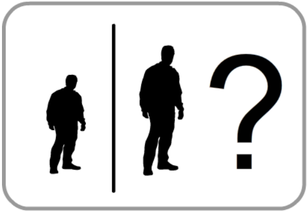
2. Probability and Bayesian Networks
   - Probability Theory
   - Bayesian Networks
   - Beyond BNs: FOL of probability

What is the problem?

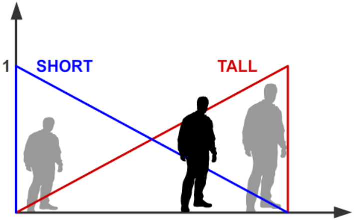
- Commonsense knowledge is not always categorical.
  - $\forall x(Bird(x) \implies Flies(x))$ fails to capture much of what we could think and/or say about the world.
- ...the world is non-categorical!
  - Statistical information.
  - Personal belief.
  - Imperfect measure tools.
  - Unreliable information.
  - Fuzzy concepts (vagueness).
  - ...

Uncertainty $\neq$ Vagueness

- John is probably tall vs. John is quite tall.

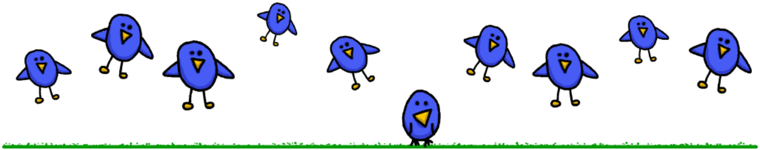


vs.



Different flavors of probability

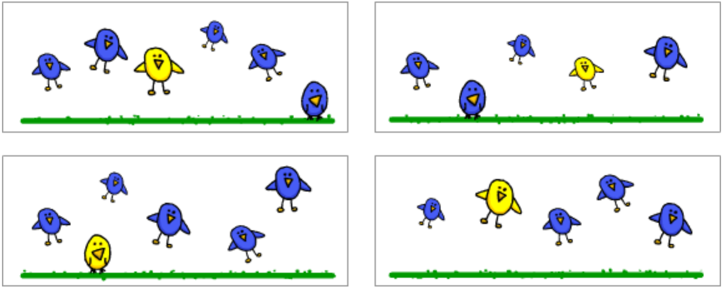
- Statistical information about the domain.
  - Things are *usually* (*rarely, almost always, ...*) a certain way.
  - "90% of birds fly"
  - $\forall x(\mathbf{P}(Bird(x) \implies Flies(x)) = 0.9)$



UIC Department of Computer Science
University of Illinois at Chicago College of Engineering

Different flavors of probability (Cont'd)

- Degree of belief.
  - Laziness, ignorance.
  - $\mathbf{P}(Flies(Tweety)) = 0.75$



  - $\mathbf{P}(\forall x(Bird(x) \implies Flies(x))) = 0.6$

From statistics to beliefs

- Going from statistical information to a degree of belief.
  - e.g. "90% of birds fly" $\longrightarrow \mathbf{P}(Flies(Tweety)) = 0.9$
- Usual procedure (*Direct Inference*):
  - Find a reference class.
  - Use statistics for that class to compute degree of belief.
- Problem: multiple reference classes.
  - Example: what is the probability that Eric is tall?
    - A 20% of American males are tall.
    - B 25% of Californian males are tall.
    - C Eric is from California.
    - D 13% of computer scientists are tall.

Objective vs. Subjective probability

A *philosophical* debate

- Objectivistic view
  - Probabilities are real aspects of the universe, propensities of objects to be a certain way.
  - Independent on who is assessing the probability.
  - Philosophical position supporting frequentist statistics.
- Subjectivistic view
  - Probability is the degree of belief of the observer, no physical significance.
  - Philosophical position supporting Bayesian statistics.
- In the end, this distinction has little practical significance.

Preliminary Concepts

- Sample space U: every possible outcome
- Events $a_i \subset U$
- $\mathcal{B}$ set of all possible events
- Probability function $\Pr : \mathcal{B} \to [0, 1]$

- Axioms of probability
  0. $\Pr(a) \geq 0$
  1. $\Pr(U) = 1$
  2. If $a_1, ..., a_n$ are disjoint events, then $\Pr(a_1 \cup ... \cup a_n) = \sum_{i=0}^{n} \Pr(a_i)$
- It follows
  3. $\Pr(\overline{a}) = 1 - \Pr(a)$
  4. $\Pr(\varnothing) = 0$

Conditioning and Bayes' Rule

1. Conditional probability: $\Pr(a|b) = \frac{\Pr(a \cap b)}{\Pr(b)}$
2. Conditional independence: $\Pr(a|s) = \Pr(a|s, b)$
   - Independece: $S = \varnothing$
3. Conjunction: $\Pr(ab) = \Pr(a|b)\Pr(b)$
4. If $\{b_1, ..., b_n\}$ is a partitioning of U then $\Pr(a) = \sum_{i=0}^{n} \Pr(a|b_i)\Pr(b_i)$
5. Bayes' Rule

$$\Pr(a|b) = \frac{\Pr(b|a)\Pr(a)}{\Pr(b)}$$

UIC Department of Computer Science
COLLEGE OF ENGINEERING

Random Variables and Joint Probability

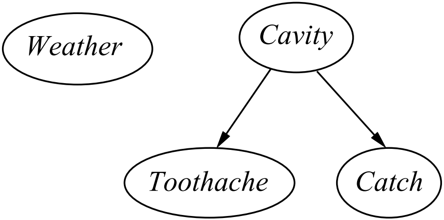
- (Propositional) random variable (r.v.): "feature" of the world whose value is uncertain.
  - e.g. X_1 outcome of the first coin toss.
  - Might be discrete or continue.
- Interpretation $\mathcal{I} \in U$: specification of the value for every r.v.
- Joint probability distribution $J(\mathcal{I})$: degree of belief the agent assigns to interpretation $\mathcal{I}$.

$$0 \leq J(\mathcal{I}) \leq 1 \quad \text{and} \quad \sum_{\mathcal{I}} J(\mathcal{I}) = 1$$

- For any event a, $\Pr(a) = \sum_{\mathcal{I} \models a} J(\mathcal{I})$
- Not useful for calculation: exponential number of interpretations.

Bayesian Networks

- A Bayesian (or belief) Network (BN) is a direct acyclic graph where:
  - **nodes** P_i are r.v.s
  - **arcs** represent (intuitively) direct dependency relations
  - each node has a **conditional probability distribution** $\Pr(P_i|Parents(P_i))$

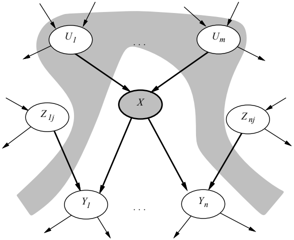


(from AIMA, 2ed)

Bayesian Networks

Basic property

- *A node is conditionally independent from its non-descendants, given its parents.*



(from AIMA, 2ed)

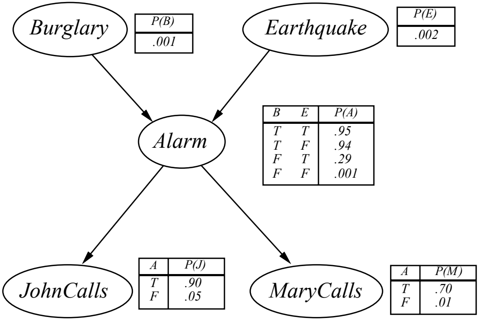
- Full joint distribution.

$$\Pr(\mathcal{I}) = \Pr(p_1, ..., p_n) = \prod_{i=1}^{n} \Pr(p_i | Parents(P_i))$$

- Usually requires much less space.

Bayesian Networks

An example



(from AIMA, 2ed)

$$
\begin{aligned}
\Pr(j \wedge m \wedge a \wedge \neg b \wedge \neg e) &= \Pr(j|a)\Pr(m|a)\Pr(a|\neg b \wedge \neg e)\Pr(\neg b)\Pr(\neg e) \\
&= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 \\
&= 0.00062
\end{aligned}
$$

UIC Department of Computer Science, College of Engineering, University of Illinois at Chicago

Queries in BNs

- **Query** variable $X \in \mathbf{X}$
- Set of **evidence** variables $\mathbf{E}$
- Set of **nonevidence** variables $\mathbf{Y} = \mathbf{X} \setminus (X \cup \mathbf{E})$

$$\Pr(X|\mathbf{e}) = \alpha \Pr(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \Pr(X, \mathbf{e}, \mathbf{y})$$

Example

$$\begin{aligned} \Pr(b|j,m) &= \alpha \sum_{e} \sum_{a} \Pr(b,e,a,j,m) \\ &= \alpha \sum_{e} \sum_{a} \Pr(b) \Pr(e) \Pr(a|b,e) \Pr(j|a) \Pr(m|a) \end{aligned}$$

UIC Department of Computer Science

Computational complexity of BN inference

- Naive computation: $O(n2^n)$
- Depth-first computation: $O(2^n)$
- Some terms are recurrent
  - Dynamic programming approach (**Variable Elimination**)
  - Linear time for polytrees (at most one path between any two nodes)
  - Exponential complexity in general
  - **Join tree** algorithms are usually used in commercial tools.
- Not surprising: BN $\supset$ Propositional Logic
  - Inference is #P-hard
- Need for approximate techniques
  - Many forms of sampling algorithms.

Beyond BNs

- BNs are essentially **propositional**.
  - Finite and fixed set of variables, with a fixed domain.
  - Don't capture regularities of the domain.
- Extend probability to First-Order
  - First-Order Probabilistic Languages (FOPL).
  - Theoretical difficulties [Halpern 1990].
  - Need to restrict the semantics.
  - Several approaches
    - Bayesian Logic (BLOG) [Milch et al.]
    - Markov Logic Networks [Domingos et al.]

UIC Department of Computer Science
COLLEGE OF ENGINEERING