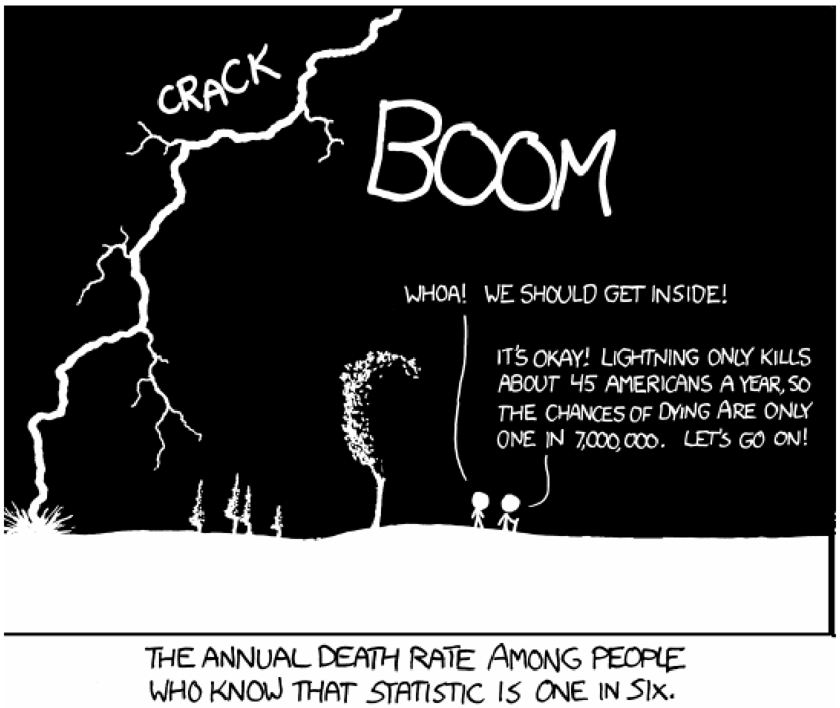
CSE4403 3.0/CSE6602E - Soft Computing
Winter 2011



Lecture 11

Probabilistic Reasoning: Why

Guest lecturer: Xiangdong An
xan@cs.yorku.ca

Reasoning

- A.k.a. inference, is the process of drawing conclusions from knowledge (assumptions or premises)
  - Logic describe rules by which reasoning operates, so that orderly reasoning can be made
- Deductive reasoning
  - Classic logic (by Aristotle based on syllogism)
- Inductive reasoning
  - Inductive logic (informal logic or critical thinking)
  - Abductive reasoning
- Formal logic (symbolic logic) – an area of mathematical logic
  - Propositional logic
  - First-order predicate logic)

Deductive reasoning

- A.k.a. deductive logic (deduction), argues from the general to a specific instance
- Syllogism

All men are mortal
Socrates is a man
Therefore, Socrates is mortal

Inductive reasoning

- A.k.a. deductive logic (deduction), argues from the particular to the general

All observed crows are black.
Therefore all crows are black.

Uncertain reasoning

- In uncertain domains, the knowledge may be vague, imprecise, incomplete or even contradictory
  - Abductive reasoning is inference from the observations to the best explanation
  - It involves reasoning in both directions

"Fire causes smoke."

Predictive reasoning: if there is fire, then there is smoke

Diagnostic reasoning: that smoke is found makes fire more credible

Abductive reasoning

- If diagnostic reasoning is ignored, some counterintuitive and strange results may appear
  - E.g., finding smoke will not make fire to be derived
- "Explaining away"
  - Both hunger and weak health can make one dizzy and finding that he has not eaten for all day makes the other reason (weak health) less credible
- Non-monotonic (non-incremental)
  - Knowledge in uncertain domains can be retracted
- Non-modularity
  - Too many exceptions and conditions (uncertainty)

Formalisms to Uncertain Reasoning

- Nonmonotonic logic
  - Extensions of classical logic FOPL with non-numerical mechanisms to deal with uncertainty
- Fuzzy logic
- Certainty factors
- Dempster-Shafer calculus
- Probabilistic reasoning
  - Bayesian probability theory

Different caculi proposed to deal with uncertainty

Unintuitive (unreasonable) results from Fuzzy logic:

T(Tall(Tom))= 0.7
T(Tall(Tom) ∧ ¬Tall(Tom))= 0.3

Probability theory

- Dates back to correspondence Fermat & Pascal 1654
  - Bayes' theorem, $P(a|b) = P(b|a)P(a)/P(b)$, shows the relation between two conditional probabilities which are the reverse of each other (by Thomas Bayes 1702-1761)
- “Probability theory is nothing but common sense reduced to calculation” (Pierre Laplace)
- “The true logic for this world is the calculus of probabilities, which takes account of the magnitude of the probability which is, or ought to be, in a reasonable man’s mind” (James Maxwell)

Bayesian probability

- Probability is held to represent the degree of plausibility of a particular statement (degree of belief)
  - A.k.a. subjective probability, epistemic probability
  - In contrast to frequentist probability (objective probability)
    - Probability is held to be derived from observed or imagined frequency distributions
- Bayes' theorem is valid regardless of whether one adopts a frequentist or a Bayesian interpretation of probability

Probabilistic reasoning

- "Coherent" iff satisfing Kolmogorov's axioms

  For any events a, b, c in sample space S (a and b are exclusive)

  $0 \leq P(a), P(b), P(c) \leq 1$ (1)
  $P(S) = 1$ (2)
  $P(a \vee b) = P(a) + P(b)$ (3)
  $P(a, c) = P(a|c)P(c)$ (1.4) (4)

- The fourth axiom introduces conditional probability:
  $P(a|c) = P(a, c)/P(c)$

- Formalized by Andrey Kolmogorov (1933)

Probabilistic reasoning

- Bayes' theorem (rule): $P(a|b) = P(b|a)P(a)/P(b)$
- Bayes' theorem to evaluate the probability of a cause given an effect:

  $P(flu|headache) = P(headache|flu)P(flu)/P(headache)$

  Why not $P(flu|headache) = P(headache, flu)/P(headache)$?

- Diagnostic knowledge is often more tenuous than the causal
  - For example, doctors do not have good idea on how many patients having headaches have flu, i.e. P(headache, flu)
  - But they do know how many patients with flu have headaches, i.e. P(headache|flu)

Probabilistic reasoning

- A joint probability distribution (JPD) specifies the probability of every elementary event in the domain
  - An elementary event is an outcome is of a sample space
  - For example, {HTT} if a coin is tossed 3 times
- Conditional representation of knowledge is more compatible with the organization of human knowledge

Reasoning based on JPD

- Based on an explicitly specified JPD, a probability distribution for any particular statement can always be computed based on Kolmogorov's 4 axioms
- However, in practice, the acquisition and updating of JPD are intractable

Acquisition intractability

- To specify a JPD $P(V)$ of n variables needs to acquire $O(d^n)$ probability values, where d is maximum size of a variable space

A JPD over V={a,b}

a	b	P (a,b)
0	0	0.1
0	1	0.2
1	0	0.3
1	1	0.4

The table grows exponentially in the number of variables

Updating intractability

- To get P(b|a=0) from a JPD,

$$P(b|a=0) = P(a=0,b) \ / \ P(a=0)$$
$$= (0.1/0.3,\ 0.2/0.3) = (1/3,\ 2/3)$$

a	b	P (a,b)
0	0	0.1
0	1	0.2
1	0	0.3
1	1	0.4

⇨

a	b	P (a,b)	P (a=0, b)
0	0	0.1	1/3
0	1	0.2	2/3
1	0	0	0
1	1	0	0

What about if you have a table with more variables, e.g. P(b,c,d,e|a=0)?

Updating intractability

Size increases exponentially (over columns a–e)

a	b	c	d	e	P(V)
0	0	0	0	0	0.01
0	0	0	0	1	0.04
0	0	0	1	0	0.1
0	0	0	1	1	0.05
…	…	…	…	…	…
1	1	1	1	0	0.14
1	1	1	1	1	0.11

Therefore, updating intractable

Marginalization intractability

- $P(a=0) = \sum_{b,c,d,e} P(a=0, b, c, d, e)$

16 rows (rows with a = 0)

16 rows (rows with a = 1)

a	b	c	d	e	P(V)
0	0	0	0	0	0.01
0	0	0	0	1	0.04
…	…	…	…	…	…
0	1	1	1	1	0.05
1	0	0	0	0	0.02
…	…	…	…	…	…
1	1	1	1	0	0.14
1	1	1	1	1	0.11

Sum half table to get P(a=0)

Marginalization intractability

Becomes half size

- $P(b,c,d,e) = \sum_a P(a,b,c,d,e)$

a	b	c	d	e	P(V)
0	0	0	0	0	0.01
0	0	0	0	1	0.04
…	…	…	…	…	…
0	1	1	1	1	0.05
1	0	0	0	0	0.02
…	…	…	…	…	…
1	1	1	1	0	0.14
1	1	1	1	1	0.11

…

⇨

b	c	d	e	P(V’)
0	0	0	0	0.03
0	0	0	1	0.14
…	…	…	…	…
1	1	1	1	0.06
0	0	0	0	0.05
…	…	…	…	…
1	1	1	0	0.18
1	1	1	1	0.16

Independency

- Belief updating based on an explicitly specified JPD assumes that everything is dependent on everything else
  - This is the reason for intractability
- Using independency can decompose problems into small problems
  - Divide and conquer
  - The problems should have structures to be solved more efficiently; otherwise, nothing could be done

Graphical models

- Graphical models make independencies in a problem explicit
  - In a graphical model, only relevant variables are adjacent
  - What one does not see locally does not matter
- Indeed, humans reason mostly by forming this kind of graphical models
- Bayesian networks (BNs), also called belief networks, causal nets are graphical models for probabilistic inference
  - Probabilistic reasoning with Bayesian networks is NP-hard (Cooper 1990)