Knowledge Engineering

Semester 2, 2004-05

Michael Rovatsos
mrovatso@inf.ed.ac.uk

Lecture 8 – Dealing with Uncertainty
8th February 2005

Where are we?

Last time . . .

- Model-based reasoning

Today . . .

- Approaches to dealing with uncertainty
  - Probabilistic Reasoning
  - Fuzzy Logic
  - Dempster-Shafer Theory

Reasoning under Uncertainty

- So far, focus on certain knowledge
  How do we model what we know?
- But how do we model uncertainty?
- Different aspects:
  - Uncertainty regarding truthfulness of propositions
  - Vagueness in the way knowledge is captured
  - Questions of ignorance and confidence
- Different KR & R approaches for each of these

Probabilistic Reasoning

- Most general and widespread method of uncertainty reasoning
- Rests on mathematical foundations of probability theory
- Two interpretations of probability:
  - Subjective: belief about likelihood of a proposition
  - Objective: frequency of observed events in which proposition holds
- Major advances in 90s, today highly popular field in AI
- Here: only very short overview (see PMR, LFD and similar courses)

Probability Theory

- Axioms of probability theory: $P(false) = 0$, $P(true) = 1$
  $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$
- Other properties: $P(\neg a) = 1 - P(a)$, $P(a) + P(\neg a) = 1$
- For discrete random variable D with domain $\langle d_1, \ldots, d_n \rangle$:
  $\sum_{i=1}^{n} P(D = d_i) = 1$
- For atomic mutually exclusive events E such that a holds in $E(a) \subseteq E$: $P(a) = \sum_{e \in E(a)} P(e)$
- Bayes' rule: $P(b|a) = \frac{P(a|b)P(b)}{P(a)}$
- Conditional independence: $P(a, b|c) = P(a|c)P(a|c)$

Example

Why is Bayes' rule useful? Assume m denotes "patient has meningitis", s denotes "patient has a stiff neck" and we have the following estimates:

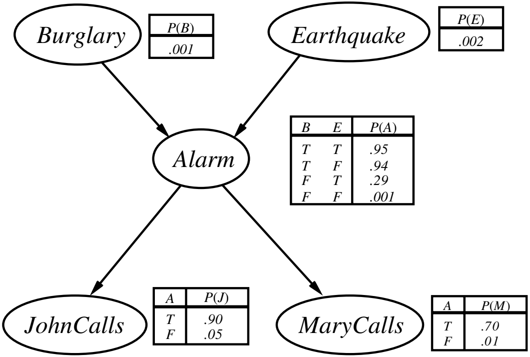
$$P(s|m) = 0.5$$
$$P(m) = 1/50000$$
$$P(s) = 1/20$$

We can infer:

$$P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{0.5 \times 1/50000}{1/20} = 0.0002$$

Belief Networks

Using a graphical notation to represent probabilities of propositions and conditional independence assumptions:



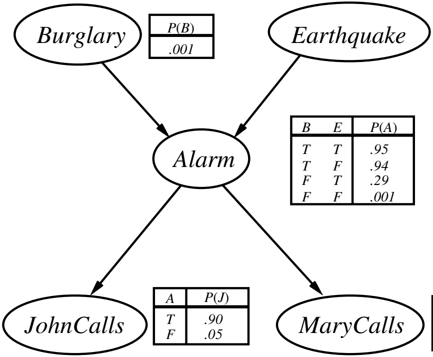
Belief Networks

- Nodes represent propositionsm, annotated with conditional probability tables
- Edges represent conditional dependencies
- Main idea: represent full joint probability distribution over variables $X_1, \ldots X_n$ (to obtain probability of conjunction $P(X_1 = x_1 \wedge \ldots \wedge X_n = x_n)$) as product of independent probabilities using Bayes' Rule
- If $parents(X_i)$ are the parent nodes of X_i, the joint probability distribution is given by

$$P(x_1, \ldots x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i))$$

Example

$$P(j \wedge m \wedge a \wedge \neg b \wedge \neg e) = P(j|a)P(m|a)P(a|\neg b \neg e)P(\neg b)P(\neg e) =$$
$$0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 = 0.00062$$



P(B)
.001

P(E)
.002

B	E	P(A)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J)
T	.90
F	.05

A	P(M)
T	.70
F	.01

Critique

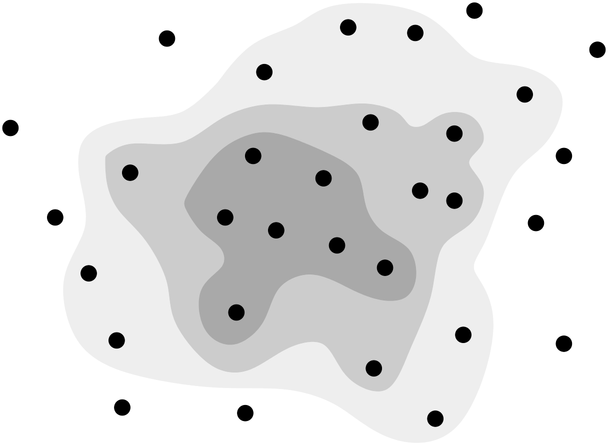
- Lots of methods for exact and approximate inference
- Area of Bayesian Learning
- Where do these probabilities come from?
- Where do the independence assumptions come from?
- Worst case: all variables depend on each other ➡ no gain

Fuzzy Logic

- Method for expressing vagueness
- Uncertainty about degree of appropriateness of a statement, not about its truthfulness
- Foundation: notion of **fuzzy sets**
- Allows for expressing degree with which an object belongs to a set and applying mathematical methods to manipulate these statements
- Fuzzy control: extremely successful in industrial applications

Fuzzy Sets

Fuzzy sets (unlike crisp ones) based on notion of “degree”



Fuzzy sets

- When describing concepts using subset relationships crisp membership often inflexible
- **Characteristic function**: members have value 1, non-members have value 0
- Take example of "young person" in terms of age
  - Naive definition: Use, for example [0,20] as a crisp interval
  - Is someone one day after his 20th birthday *not young*?
  - Note that this problem appears regardless of the bound
- Solution: Allow more intermediate values for characteristic function (gradual membership)

Operations on fuzzy sets

- Logical operations define
- Let $T(A)$, $T(B)$ fuzzy truth values of A and B
  - $T(A \wedge B) = \min(T(A), T(B))$
  - $T(A \vee B) = \max(T(A), T(B))$
  - $T(\neg A) = 1 - T(A)$
- **Truth-functional approach** ➡ problems with correlations and anti-correlations between propositions
- Example: Fuzzy truth value of "tall and heavy" will be unreasonably high for someone who is extremely tall although "heavy" should be less strict for very tall people

Further Issues

- Devise rules of the form "if person is young and not overweight then blood pressure is normal" to make decisions
- **Defuzzification**: how to make crisp choices after evaluation of fuzzy rules (e.g. take center of gravity of a fuzzy set)
- Attempts to map fuzzy logic to probabilistic concepts
  - Discrete observation interpretation: $P(\textit{Observer says person is tall and heavy} \mid \textit{Height}, \textit{Weight})$
    ➡ solves truth-functionality problems
  - Random set interpretation: view *Tall* as a random variable (denoting a set), $P(\textit{Tall} = S)$ is probability that set S of persons would be identified as tall

Critique

- Success in practical applications often attributed to:
  - Use in limited, controllable domains
  - Fine-tuning of parameters for a particular use
  - No chaining of inferences
- Hard to combine with other kinds of KBS
- However, so far the only AI technology that has found its way to (almost) every washing machine!

Dempster-Shafer Theory

- Based on dealing with distinction between **uncertainty** and **ignorance**
- Computes probability that evidence supports proposition (rather than probability that proposition is true)
- Two elements:
  - Obtaining degrees of belief for one question from subjective probabilities for related question
  - Combining such degrees of belief when they are based on independent items of evidence

Example

- Betty is reliable with probability 0.9
- She says a limb fell on my car (proposition A)
  - A is not necessarily false if she is unreliable
  - Statement justifies a degree of belief of 0.9 in A, and zero degree of belief (not 0.1) that $\neg A$
  - This does not mean I am sure $\neg A$ is not the case, but that have no evidence to believe otherwise
- Suppose Sally is also reliable with probability 0.9 and she also claims A
  - Probability of both being reliable is 0.81, and of at least one being reliable is 1-0.01=0.99 ➡ my degree of belief in A is 0.99

Example

- Suppose they contradict each other (Sally says $\neg A$):
  - Probabilities that only Betty/only Sally/neither of them is reliable are 0.09/0.09/0.01, normalised 9/19, 9/19, 1/19
  - Belief of 9/19 that A and belief of 9/19 that $\neg A$
- Begin with assumption that two questions (Did limb fall on car? Is the witness reliable?) are independent
- Independence disappears when conflict between different items of evidence becomes apparent

Critique

- Strength of DS theory: discriminating between ignorance and uncertainty
- Ease of representation of evidence at different levels of abstraction
- "Interval" view appealing
  - In our example, before evidence probability of A can be from [0,1]
  - After evidence [0.99,1] (if they agree) [9/19,10/19] (if they disagree)
- However, in a complete Bayesian model evidence can be included as a variable

Summary

- Overview of different aspects of uncertainty
  - Probabilistic approach: assigning probabilities to truth value of propositions
  - Fuzzy logic approach: assessing how appropriate a proposition is under certain properties of an object
  - Dempster-Shafer theory: assigning degrees of belief vs. ignorance given some evidence
  - (Default reasoning)
- Completes our account of knowledge representation and reasoning
- Next block: **Knowledge Synthesis**
- Next lecture: Automated software synthesis