Uncertainty in Expert Systems

Based on Chapter 3 in Negnevitsky

Sources of uncertainty

- **Weak implications**: Want to be able to capture associations and correlations, not just cause and effect.
- **Imprecise language**: How often is “sometimes”?
- **Unknown data**: In real problems, data is often incomplete or missing
- **Differing experts**: Experts often disagree, or have different reasons for agreeing.

Approach 1: Bayesian reasoning

Probability of independent outcomes

If s is the number of times success can occur, and f is the number of times failure can occur:

$$P(\text{success}) = p = \frac{s}{s+f}$$

$$P(\text{failure}) = q = \frac{f}{s+f}$$

$$p + q = 1$$

Conditional probability

The probability of A given B is:

(1)

$$p(A \mid B) = \frac{p(A \cap B)}{p(B)}$$

Similarly, for B given A:

$$p(B \mid A) = \frac{p(B \cap A)}{p(A)}$$

Conditional probability (2)

Joint probability is commutative, so:

$$p(A \cap B) = p(B \mid A) * p(A)$$

Substitute into (1) to get:

$$p(A \mid B) = \frac{p(B \mid A) * p(A)}{p(B)}$$

This is the ***Bayesian Rule***.

Conditional probability (3)

If A depends on many ***mutually exclusive*** events:

(2) $$\sum_{i=1}^{n} p(A \cap B) = \sum_{i=1}^{n} p(A \mid B_i) * p(B_i)$$

If ***all*** events with which A can occur are included, then:

$$\sum_{i=1}^{n} p(A \cap B_i) = p(A)$$

Conditional probability (4)

So (2) simplifies to:

(3) $$p(A) = \sum_{i=1}^{n} p(A \mid B_i) * p(B_i)$$

Conditional probability (5)

If A depends only on B and not B, then (3) becomes:

$$p(A) = p(A \mid B) * p(B) + p(A \mid \neg B) * p(\neg B)$$

If you do the same for p(B), and substitute into the Bayesian rule:

$$p(A \mid B) = \frac{p(B \mid A) * p(A)}{p(B \mid A) * p(A) + p(B \mid \neg A) * p(\neg A)}$$

(4)

Bayesian reasoning

- Instead of A and B, consider H (a hypothesis) and E (evidence for that hypothesis).
- Given E, what is the probability of H being true?
- P(H|E) is referred to as the ***posterior probability*** of H given evidence E.
- Expert provides prior probabilities for p(H), p(not H), as well as conditional probabilities p(E|H) and p(E|not H).

Bayesian reasoning: multiple hypotheses and evidences

For many hypotheses and evidences, we can generalize (4):

$$p(H_i \mid E_1 \ldots E_n) = \frac{p(E_1 \ldots E_n \mid H_i) * p(H_i)}{\sum_{k=1}^{n} p(E_1 \ldots E_n \mid H_k) * p(H_k)}$$

However, this is very awkward to use – requires conditional probabilities for all possible combos of evidences and hypotheses!!

Bayesian reasoning: multiple hypotheses and evidences (2)

If we assume conditional independence among different evidences, simplifies to:

$$p(H_i \mid E_1 \ldots E_n) = \frac{p(E_1 \mid H_i) * \ldots * p(E_n \mid H_i) * p(H_i)}{\sum_{k=1}^{m} p(E_1 \mid H_k) * \ldots * p(E_n \mid H_k) * p(H_k)}$$

By "conditional independence", we mean:

$$p(E_1 \ldots E_n \mid H) = p(E_1 \mid H) * \ldots * p(E_n \mid H)$$

In class exercise

Your expert gives you the following table of prior and conditional probabilities:

	Hypothesis		
Probability	i=1	i=2	i=3
$p(H_i)$	0.25	0.35	0.4
$P(E_1\|H_i)$	0.6	0.2	0.3
$P(E_2\|H_i)$	0.0	0.5	0.7
$P(E_3\|H_i)$	0.4	0.3	0.25

The user observes E_2 and E_3 – what is the most likely hypothesis?

Likelihood of sufficiency

Measure of an expert's level of belief in H if E is present.

$$LS = \frac{p(E \mid H)}{p(E \mid \neg H)}$$

More intuitive for an expert than conditional probabilities.

Likelihood of necessity

Measure of an expert's disbelief in H if E is not present:

$$LN = \frac{p(\neg E \mid H)}{p(\neg E \mid \neg H)}$$

LN and LS cannot be derived from each other – the expert (or a statistical analysis of a dataset!) must provide both.

Prior odds

For simplicity of calculation, we use *prior odds* (instead of prior probability):

$$O(H) = \frac{p(H)}{1 - p(H)} \qquad (5)$$

If the rule antecedent (i.e. the evidence) is true:

$$O(H \mid E) = LS * O(H)$$

If it is false:

$$O(H \mid \neg E) = LN * O(H)$$

Posterior probabilities can be recovered using (5).

Problems with the Bayesian approach

- Humans are not very good at estimating probability!
  - In particular, we tend to make different assumptions when calculating prior and conditional probabilities
- Reliable and complete statistical information often not available.
- Bayesian approach requires hypotheses to be independent and exhaustive, and evidences to be at least conditionally independent – often not the case.
- One solution: certainty factors

Certainty factors

Based on Ch 3 in Negnevitsky

When is classical probability not applicable?

- Experts unable to express strength of belief in mathematically consistent way
- Reliable statistical data not available
- Independence of hypotheses and/or evidences not clear

Certainty factor (CF)

- A measure of an expert's belief in a fact or rule.
- Ranges from –1 to 1.
- "Fuzzy" reasoning about probability.

Terms (examples)	CF
Definitely not	-1.0
Almost certainly not	-0.8
Probably not	-0.6
Maybe not	-0.4
Unknown	-0.2-0.2
Maybe	0.4
Probably	0.6
Almost certainly	0.8
Definitely	1.0

CFs, MD and MB

CFs are related to probabilities via two functions: measure of belief MB(H,E) and measure of disbelief MD(H,E).

- MB(H,E): the degree to which belief in H is increased if E is observed.
- MD(H,E): the degree to which belief in H is decreased if E is observed.

CFs, MD and MB (2)

MD and MB are related to probability by the following equations:

$$MB(H,E) = \begin{cases} 1 & \text{If p(H)=1} \\ \dfrac{\max[\,p(H \mid E), p(H)] - p(H)}{1 - p(H)} & \text{otherwise} \end{cases}$$

$$MD(H,E) = \begin{cases} 1 & \text{If p(H)=0} \\ \dfrac{\min[\,p(H \mid E), p(H)] - p(H)}{-\,p(H)} & \text{otherwise} \end{cases}$$

CFs, MD and MB (3)

MB and MD are combined into one number (CF) by this equation:

$$cf = \frac{MB(H,E) - MD(H,E)}{1 - \min[MB(H,E), MD(H,E)]}$$

Use of CFs

CFs capture the certainty that a given rule holds. For example:

- IF animal lays eggs THEN animal is bird {cf 0.3}
- IF animal lays eggs THEN animal is reptile {cf 0.2}

Note that this allows 'inconsistent' rules to co-exist in the knowledge base.

Use of CFs (2)

For conciseness, can also assign CFs to different consequents, e.g.:

- If animal lays eggs THEN
  - animal is bird {CF 0.3}
  - animal is reptile {CF 0.2}

Note that CFs don’t need to add up to 1.

Use of CFs (3)

CFs are also applied to evidence (facts), e.g.:

- animal lays eggs {cf 0.6}
- temperature is high {cf 0.9}
- user likes red wine {cf 0} (i.e. user has never tried red wine)

Represents reliability or availability of evidence. Typically given by the user at run time.

Propagation of CFs

For a single antecedent rule:

$$cf(H,E) = cf(E) * cf(R)$$

Where cf(E) is the CF of the evidence, and cf(R) is the CF of the rule.

Single antecedent rule example

- IF patient has toothache THEN problem is cavity {cf 0.3}
- Patient has toothache {cf 0.9}

What is the cf(cavity, toothache)?

$$cf = 0.9 * 0.3 = 0.27$$

i.e. "weak maybe"

Propagation of CFs (multiple antecedents)

For conjunctive rules:

$$cf(H, E_1 \cap \ldots \cap E_n) = \min[cf(E_1) \ldots cf(E_n)] * cf(R)$$

For disjunctive rules:

$$cf(H, E_1 \cup \ldots \cup E_n) = \max[cf(E_1) \ldots cf(E_n)] * cf(R)$$

Conjunctive example

- IF patient has toothache AND patient has prior cavities THEN problem is cavity {cf 0.3}
- Patient has toothache {cf -0.5}
- Patient has prior cavities {cf 0.9}
- What is the cf for "problem is cavity"?

$$cf = \min(-0.5, 0.9) * 0.3 = -0.15$$

Disjunctive example

- IF patient has toothache OR patient has prior cavities THEN problem is cavity {cf 0.3}
- Patient has toothache {cf -0.5}
- Patient has prior cavities {cf 0.9}
- What is the cf for “problem is cavity”?

$$cf = \max(-0.5, 0.9) * 0.3 = 0.27$$

Multiple rules affecting H

If the hypothesis is affected by several rules:

$$cf(cf_1, cf_2) = \begin{cases} cf_1 + cf_2 * (1 - cf_1) & \text{If cf1 and cf2 > 0} \\ \dfrac{cf_1 + cf_2}{1 - \min(|cf_1|, |cf_2|)} & \text{If cf1 xor cf2 < 0} \\ cf_1 + cf_2 * (1 + cf_1) & \text{If cf1 and cf2 < 0} \end{cases}$$

Multiple rules example:

- IF patient has toothache THEN problem is cavity {cf 0.3}
- IF patient has prior cavities THEN problem is cavity {cf 0.7}
- Patient has toothache (cf –0.5)
- Patient has prior cavities (0.9)

Calculate cf for “problem is cavity”.

Bayesian vs certainty factors

- Prob theory is ‘good math’, and works well if statistical data is available and accurate probability statements can be made
- CF theory lacks formal mathematical foundation, but better able the kind of estimates an expert is likely to make of probability

Exercise

- Add CFs to the rules in the “lights out” rule base.
- Run through an example, and calculate resulting CF