Unit 5 : Uncertainty Management in Rule Based Expert System

Prepared by Prof.Kore S.S

(Assistant Professor(CSE) DYPCET,Kolhapur)

Uncertainty in expert systems

- Uncertainty can be defined as the lack of the exact knowledge that would enable us to reach a perfectly reliable conclusion
- e.g IF A is true THEN A is not false
- e.g B is false THEN B is not true

- The available information often contains inexact, incomplete or even unmeasurable data.

Sources of uncertain knowledge

- Weak implications.

Rule-based expert systems often suffer from weak implications and vague associations. Domain experts and knowledge engineers have the painful, and rather hopeless, task of establishing concrete correlations between IF (condition) and THEN (action) parts of the rules.

- Imprecise language.

Our natural language is inherently ambiguous and imprecise. We describe facts with such terms as often and sometimes, frequently and hardly ever. As a result, it can be difficult to express knowledge in the precise IF-THEN form of production rules

Table 3.1 Quantification of ambiguous and imprecise terms on a time-frequency scale

Ray Simpson (1944)		Milton Hakel (1968)	
Term	**Mean value**	**Term**	**Mean value**
Always	99	Always	100
Very often	88	Very often	87
Usually	85	Usually	79
Often	78	Often	74
Generally	78	Rather often	74
Frequently	73	Frequently	72
Rather often	65	Generally	72
About as often as not	50	About as often as not	50
Now and then	20	Now and then	34
Sometimes	20	Sometimes	29
Occasionally	20	Occasionally	28
Once in a while	15	Once in a while	22
Not often	13	Not often	16
Usually not	10	Usually not	16
Seldom	10	Seldom	9
Hardly ever	7	Hardly ever	8
Very seldom	6	Very seldom	7
Rarely	5	Rarely	5
Almost never	3	Almost never	2
Never	0	Never	0

- Unknown data.

When the data is incomplete or missing, the only solution is to accept the value ‘unknown’ and proceed to an approximate reasoning with this value

- Combining the views of different experts.

Large expert systems usually combine the knowledge and expertise of a number of experts. Unfortunately, experts seldom reach exactly the same conclusions. Usually, experts have contradictory opinions and produce conflicting rules.

Basic probability theory

- Probability can be expressed mathematically as a numerical index with a range between zero (an absolute impossibility) to unity (an absolute certainty). Most events have a probability index strictly between 0 and 1, which means that each event has at least two possible outcomes: favorable outcome or success, and unfavorable outcome or failure.

The probability of success and failure can be determined as follows:

$$P(\text{success}) = \frac{\text{the number of successes}}{\text{the number of possible outcomes}} \tag{3.1}$$

$$P(\text{failure}) = \frac{\text{the number of failures}}{\text{the number of possible outcomes}} \tag{3.2}$$

Therefore, if s is the number of times success can occur, and f is the number of times failure can occur, then

$$P(\text{success}) = p = \frac{s}{s+f} \tag{3.3}$$

$$P(\text{failure}) = q = \frac{f}{s+f} \tag{3.4}$$

and

$$p + q = 1 \tag{3.5}$$

Let us consider classical examples with a coin and a die. If we throw a coin, the probability of getting a head will be equal to the probability of getting a tail. In a single throw, $s = f = 1$, and therefore the probability of getting a head (or a tail) is 0.5.

Consider now a dice and determine the probability of getting a 6 from a single throw. If we assume a 6 as the only success, then $s = 1$ and $f = 5$, since there is just one way of getting a 6, and there are five ways of not getting a 6 in a single throw. Therefore, the probability of getting a 6 is

$$p = \frac{1}{1+5} = 0.1666$$

and the probability of not getting a 6 is

$$q = \frac{5}{1+5} = 0.8333$$

Conditional Probability

- Let A be an event in the world and B be another event.
- The probability that event A will occur if event B occurs is called the conditional probability. Conditional probability is denoted mathematically as P(A|B) in which the vertical bar represents GIVEN and the complete probability expression is interpreted as ‘Conditional probability of event A occurring given that event B has occurred’

$$p(A|B) = \frac{\text{the number of times } A \text{ and } B \text{ can occur}}{\text{the number of times } B \text{ can occur}}$$

Joint probability

The number of times A and B can occur, or the probability that both A and B will occur, is called the **joint probability** of A and B. It is represented mathematically as $p(A \cap B)$. The number of ways B can occur is the probability of B, $p(B)$, and thus

$$p(A|B) = \frac{p(A \cap B)}{p(B)} \tag{3.7}$$

Similarly, the conditional probability of event B occurring given that event A has occurred equals

$$p(B|A) = \frac{p(B \cap A)}{p(A)} \tag{3.8}$$

Hence,

$$p(B \cap A) = p(B|A) \times p(A) \tag{3.9}$$

The joint probability is commutative, thus

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

Therefore,

$$p(A \cap B) = p(B|A) \times p(A) \tag{3.10}$$

Substituting Eq. (3.10) into Eq. (3.7) yields the following equation:

$$p(A|B) = \frac{p(B|A) \times p(A)}{p(B)}, \tag{3.11}$$

where:

$p(A|B)$ is the conditional probability that event A occurs given that event B has occurred;

$p(B|A)$ is the conditional probability of event B occurring given that event A has occurred;

$p(A)$ is the probability of event A occurring;

$p(B)$ is the probability of event B occurring.

Equation (3.11) is known as the **Bayesian rule**, which is named after Thomas

Bayesian reasoning

Suppose all rules in the knowledge base are represented in the following

IF E is true

THEN H is true {with probability p}

This rule implies that if event E occurs, then the probability that event H will occur is p.

- We simply use H and E instead of A and B. In expert systems, H usually represents a hypothesis and E denotes evidence to support this hypothesis

$$p(H|E) = \frac{p(E|H) \times p(H)}{p(E|H) \times p(H) + p(E|\neg H) \times p(\neg H)} \tag{3.18}$$

where:

$p(H)$ is the prior probability of hypothesis H being true;

$p(E|H)$ is the probability that hypothesis H being true will result in evidence E;

$p(\neg H)$ is the prior probability of hypothesis H being false;

$p(E|\neg H)$ is the probability of finding evidence E even when hypothesis H is false.

Probability p(H|E) is called the posterior probability of hypothesis H upon observing evidence E

What if the expert, based on single evidence E, cannot choose a single hypothesis but rather provides multiple hypotheses $H_1, H_2, \ldots, H_m$? Or given multiple evidences $E_1, E_2, \ldots, E_n$, the expert also produces multiple hypotheses?

We can generalise Eq. (3.18) to take into account both multiple hypotheses $H_1, H_2, \ldots, H_m$ and multiple evidences $E_1, E_2, \ldots, E_n$. But the hypotheses as well as the evidences must be mutually exclusive and exhaustive.

Single evidence E and multiple hypotheses $H_1, H_2, \ldots, H_m$ follow:

$$p(H_i|E) = \frac{p(E|H_i) \times p(H_i)}{\sum_{k=1}^{m} p(E|H_k) \times p(H_k)} \tag{3.19}$$

Multiple evidences $E_1, E_2, \ldots, E_n$ and multiple hypotheses $H_1, H_2, \ldots, H_m$ follow:

$$p(H_i|E_1 E_2 \ldots E_n) = \frac{p(E_1 E_2 \ldots E_n|H_i) \times p(H_i)}{\sum_{k=1}^{m} p(E_1 E_2 \ldots E_n|H_k) \times p(H_k)} \tag{3.20}$$

$$p(H_i|E_1E_2\ldots E_n) = \frac{p(E_1|H_i) \times p(E_2|H_i) \times \ldots \times p(E_n|H_i) \times p(H_i)}{\sum_{k=1}^{m} p(E_1|H_k) \times p(E_2|H_k) \times \ldots \times p(E_n|H_k) \times p(H_k)} \qquad (3.21)$$

How does an expert system compute all posterior probabilities and finally rank potentially true hypotheses?

Table 3.2 illustrates the prior and conditional probabilities provided by the expert.

Assume that we first observe evidence E_3. The expert system computes the posterior probabilities for all hypotheses according to Eq. (3.19):

$$p(H_i|E_3) = \frac{p(E_3|H_i) \times p(H_i)}{\sum_{k=1}^{3} p(E_3|H_k) \times p(H_k)}, \qquad i = 1, 2, 3$$

Thus,

$$p(H_1|E_3) = \frac{0.6 \times 0.40}{0.6 \times 0.40 + 0.7 \times 0.35 + 0.9 \times 0.25} = 0.34$$

$$p(H_2|E_3) = \frac{0.7 \times 0.35}{0.6 \times 0.40 + 0.7 \times 0.35 + 0.9 \times 0.25} = 0.34$$

$$p(H_3|E_3) = \frac{0.9 \times 0.25}{0.6 \times 0.40 + 0.7 \times 0.35 + 0.9 \times 0.25} = 0.32$$

Table 3.2 The prior and conditional probabilities

	Hypothesis		
Probability	$i = 1$	$i = 2$	$i = 3$
$p(H_i)$	0.40	0.35	0.25
$p(E_1\|H_i)$	0.3	0.8	0.5
$p(E_2\|H_i)$	0.9	0.0	0.7
$p(E_3\|H_i)$	0.6	0.7	0.9

As you can see, after evidence E_3 is observed, belief in hypothesis H_1 decreases and becomes equal to belief in hypothesis H_2. Belief in hypothesis H_3 increases and even nearly reaches beliefs in hypotheses H_1 and H_2.

Suppose now that we observe evidence E_1. The posterior probabilities are calculated by Eq. (3.21):

$$p(H_i|E_1E_3) = \frac{p(E_1|H_i) \times p(E_3|H_i) \times p(H_i)}{\sum_{k=1}^{3} p(E_1|H_k) \times p(E_3|H_k) \times p(H_k)}, \qquad i = 1, 2, 3$$

Hence,

$$p(H_1|E_1E_3) = \frac{0.3 \times 0.6 \times 0.40}{0.3 \times 0.6 \times 0.40 + 0.8 \times 0.7 \times 0.35 + 0.5 \times 0.9 \times 0.25} = 0.19$$

$$p(H_2|E_1E_3) = \frac{0.8 \times 0.7 \times 0.35}{0.3 \times 0.6 \times 0.40 + 0.8 \times 0.7 \times 0.35 + 0.5 \times 0.9 \times 0.25} = 0.52$$

$$p(H_3|E_1E_3) = \frac{0.5 \times 0.9 \times 0.25}{0.3 \times 0.6 \times 0.40 + 0.8 \times 0.7 \times 0.35 + 0.5 \times 0.9 \times 0.25} = 0.29$$

Hypothesis H_2 is now considered as the most likely one, while belief in hypothesis H_1 has decreased dramatically.

After observing evidence E_2 as well, the expert system calculates the final posterior probabilities for all hypotheses:

$$p(H_i|E_1E_2E_3) = \frac{p(E_1|H_i) \times p(E_2|H_i) \times p(E_3|H_i) \times p(H_i)}{\sum_{k=1}^{3} p(E_1|H_k) \times p(E_2|H_k) \times p(E_3|H_k) \times p(H_k)}, \qquad i = 1, 2, 3$$

Thus,

$$p(H_1|E_1E_2E_3) = \frac{0.3 \times 0.9 \times 0.6 \times 0.40}{0.3 \times 0.9 \times 0.6 \times 0.40 + 0.8 \times 0.0 \times 0.7 \times 0.35 + 0.5 \times 0.7 \times 0.9 \times 0.25}$$
$$= 0.45$$

$$p(H_2|E_1E_2E_3) = \frac{0.8 \times 0.0 \times 0.7 \times 0.35}{0.3 \times 0.9 \times 0.6 \times 0.40 + 0.8 \times 0.0 \times 0.7 \times 0.35 + 0.5 \times 0.7 \times 0.9 \times 0.25}$$
$$= 0$$

$$p(H_3|E_1E_2E_3) = \frac{0.5 \times 0.7 \times 0.9 \times 0.25}{0.3 \times 0.9 \times 0.6 \times 0.40 + 0.8 \times 0.0 \times 0.7 \times 0.35 + 0.5 \times 0.7 \times 0.9 \times 0.25}$$
$$= 0.55$$

- Although the initial ranking provided by the expert was H1, H2 and H3, only hypotheses H1 and H3 remain under consideration after all evidences (E1, E2 and E3) were observed. Hypothesis H2 can now be completely abandoned. Note that hypothesis H3 is considered more likely than hypothesis H1.

Table 3.4 Uncertain terms and their interpretation

Term	Certainty factor
Definitely not	–1.0
Almost certainly not	–0.8
Probably not	–0.6
Maybe not	–0.4
Unknown	–0.2 to +0.2
Maybe	+0.4
Probably	+0.6
Almost certainly	+0.8
Definitely	+1.0

certainty factor (*cf*), a number to measure the expert's belief. The maximum value of the certainty factor was +1.0 (definitely true) and the minimum –1.0 (definitely false). A positive value represented a degree of belief and a negative a degree of disbelief. For example, if the expert stated that some evidence was almost certainly true, a *cf* value of 0.8 would be assigned to this evidence. Table 3.4 shows some uncertain terms interpreted in MYCIN (Durkin, 1994).

In expert systems with certainty factors, the knowledge base consists of a set of rules that have the following syntax:

IF <evidence>
THEN <hypothesis> {*cf*}

where *cf* represents belief in hypothesis *H* given that evidence *E* has occurred.

uncertain. The net certainty for a single antecedent rule, $cf(H,E)$, can be easily calculated by multiplying the certainty factor of the antecedent, $cf(E)$, with the rule certainty factor, cf

$$cf(H,E) = cf(E) \times cf \tag{3.32}$$

For example,

IF the sky is clear
THEN the forecast is sunny {cf 0.8}

and the current certainty factor of *sky is clear* is 0.5, then

$$cf(H,E) = 0.5 \times 0.8 = 0.4$$

This result, according to Table 3.4, would read as 'It may be sunny'.

How does an expert system establish the certainty factor for rules with multiple antecedents?

For conjunctive rules such as

IF <evidence E_1>
AND <evidence E_2>
.
.
.
AND <evidence E_n>
THEN <hypothesis H> {*cf*}

the net certainty of the consequent, or in other words the certainty of hypothesis *H*, is established as follows:

$$cf(H, E_1 \cap E_2 \cap \ldots \cap E_n) = min\,[cf(E_1), cf(E_2), \ldots, cf(E_n)] \times cf \tag{3.33}$$

For example,

IF sky is clear
AND the forecast is sunny
THEN the action is 'wear sunglasses' {*cf* 0.8}

and the certainty of *sky is clear* is 0.9 and the certainty of the *forecast is sunny* is 0.7, then

$$cf(H, E_1 \cap E_2) = min\,[0.9, 0.7] \times 0.8 = 0.7 \times 0.8 = 0.56$$

According to Table 3.4, this conclusion might be interpreted as 'Probably it would be a good idea to wear sunglasses today'.

For disjunctive rules such as

For disjunctive rules such as

IF <evidence E_1>
OR <evidence E_2>
.
.
.
OR <evidence E_n>
THEN <hypothesis H> {cf}

the certainty of hypothesis H, is determined as follows:

$$cf(H, E_1 \cup E_2 \cup \ldots \cup E_n) = max\,[cf(E_1), cf(E_2), \ldots, cf(E_n)] \times cf \tag{3.34}$$

For example,

IF sky is overcast
OR the forecast is rain
THEN the action is 'take an umbrella' {cf 0.9}

and the certainty of *sky is overcast* is 0.6 and the certainty of the *forecast is rain* is 0.8, then

$$cf(H, E_1 \cup E_2) = max\,[0.6, 0.8] \times 0.9 = 0.8 \times 0.9 = 0.72,$$

which can be interpreted as 'Almost certainly an umbrella should be taken today'.

Combined Certainty Factor

- Sometimes two or even more rules can affect the same hypothesis

Rule 1:	IF	A is X
	THEN	C is Z {cf 0.8}
Rule 2:	IF	B is Y
	THEN	C is Z {cf 0.6}

What certainty should be assigned to object C having value Z if both Rule 1 and Rule 2 are fired? Our common sense suggests that, if we have two pieces of evidence (A is X and B is Y) from different sources (Rule 1 and Rule 2) supporting the same hypothesis (C is Z), then the confidence in this hypothesis should increase and become stronger than if only one piece of evidence had been obtained.

To calculate a combined certainty factor we can use the following equation (Durkin, 1994):

$$cf(cf_1, cf_2) = \begin{cases} cf_1 + cf_2 \times (1 - cf_1) & \text{if } cf_1 > 0 \text{ and } cf_2 > 0 \\ \dfrac{cf_1 + cf_2}{1 - min\,[|cf_1|, |cf_2|]} & \text{if } cf_1 < 0 \text{ or } cf_2 < 0 \\ cf_1 + cf_2 \times (1 + cf_1) & \text{if } cf_1 < 0 \text{ and } cf_2 < 0 \end{cases} \tag{3.35}$$

where:

cf_1 is the confidence in hypothesis H established by Rule 1;
cf_2 is the confidence in hypothesis H established by Rule 2;
$|cf_1|$ and $|cf_2|$ are absolute magnitudes of cf_1 and cf_2, respectively.

Thus, if we assume that

$$cf(E_1) = cf(E_2) = 1.0$$

then from Eq. (3.32) we get:

$$\begin{aligned} cf_1(H, E_1) &= cf(E_1) \times cf_1 = 1.0 \times 0.8 = 0.8 \\ cf_2(H, E_2) &= cf(E_2) \times cf_2 = 1.0 \times 0.6 = 0.6 \end{aligned}$$

and from Eq. (3.35) we obtain:

$$\begin{aligned} cf(cf_1, cf_2) &= cf_1(H, E_1) + cf_2(H, E_2) \times [1 - cf_1(H, E_1)] \\ &= 0.8 + 0.6 \times (1 - 0.8) = 0.92 \end{aligned}$$

This example shows an incremental increase of belief in a hypothesis and also confirms our expectations.

Consider now a case when rule certainty factors have the opposite signs. Suppose that

$$cf(E_1) = 1 \text{ and } cf(E_2) = -1.0,$$

then

$$cf_1(H, E_1) = 1.0 \times 0.8 = 0.8$$
$$cf_2(H, E_2) = -1.0 \times 0.6 = -0.6$$

and from Eq. (3.35) we obtain:

$$cf(cf_1, cf_2) = \frac{cf_1(H, E_1) + cf_2(H, E_2)}{1 - min\,[|cf_1(H, E_1)|, |cf_2(H, E_2)|]} = \frac{0.8 - 0.6}{1 - min\,[0.8, 0.6]} = 0.5$$

This example shows how a combined certainty factor, or in other words net belief, is obtained when one rule, Rule 1, confirms a hypothesis but another, Rule 2, discounts it.

Let us consider now the case when rule certainty factors have negative signs. Suppose that:

$$cf(E_1) = cf(E_2) = -1.0,$$

then

$$cf_1(H, E_1) = -1.0 \times 0.8 = -0.8$$
$$cf_2(H, E_2) = -1.0 \times 0.6 = -0.6$$

and from Eq. (3.35) we obtain:

$$\begin{aligned} cf(cf_1, cf_2) &= cf_1(H, E_1) + cf_2(H, E_2) \times [1 + cf_1(H, E_1)] \\ &= -0.8 - 0.6 \times (1 - 0.8) = -0.92 \end{aligned}$$

This example represents an incremental increase of disbelief in a hypothesis.