BMCS2003
Artificial Intelligence

Uncertainty in Expert System

what is uncertainty?

uncertain

incomplete

information

Unsuitable to solve a problem

inconsistent

Example 1

- IF rain clouds present
- THEN rain?? Drizzle??

Example 2

- IF passed the test THEN passed the final exam

Objectives

- Uncertainty vs. classical Expert System
- Basic probability theory
- Bayesian reasoning
- Certainty factors theory
- fuzzy logic

Uncertainty v.s. classical ES

Uncertainty

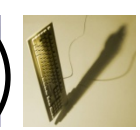
- lack of the exact knowledge that still enable us to reach a perfectly reliable conclusion

Classic logic

- permits only exact reasoning.
- *law of the excluded middle*
- A → B

TRUE FALSE

TRUE FALSE



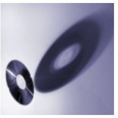
Sources of Uncertain knowledge

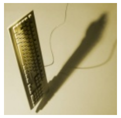
Weak implications of Classic ES

- Domain experts and knowledge engineers have the painful task of establishing concrete correlations between IF (condition) and THEN (action) parts of the rules.

Vague associations

- need to have the ability to handle vague associations, for example by accepting the degree of correlations as numerical certainty factors.





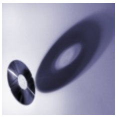
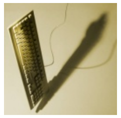
Sources of Uncertain Knowledge

Imprecise language

- natural language is ambiguous and imprecise.
- e.g. "*often*", "*sometimes*", "*frequently*", "*hardly ever*"

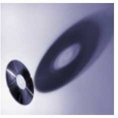
Unknown data

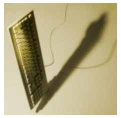
- When the data is incomplete or missing, the only solution is to accept the value "unknown" and proceed to an approximate reasoning with this value.
- e.g. "*I don't know*", "*I am not sure*", "*I don't understand*"

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

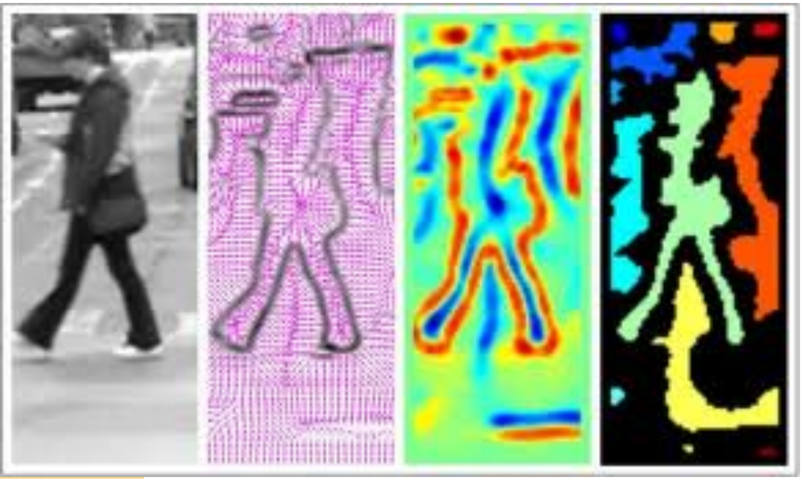


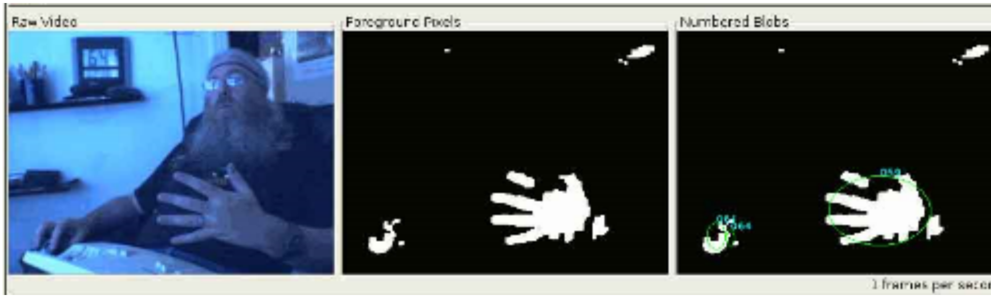


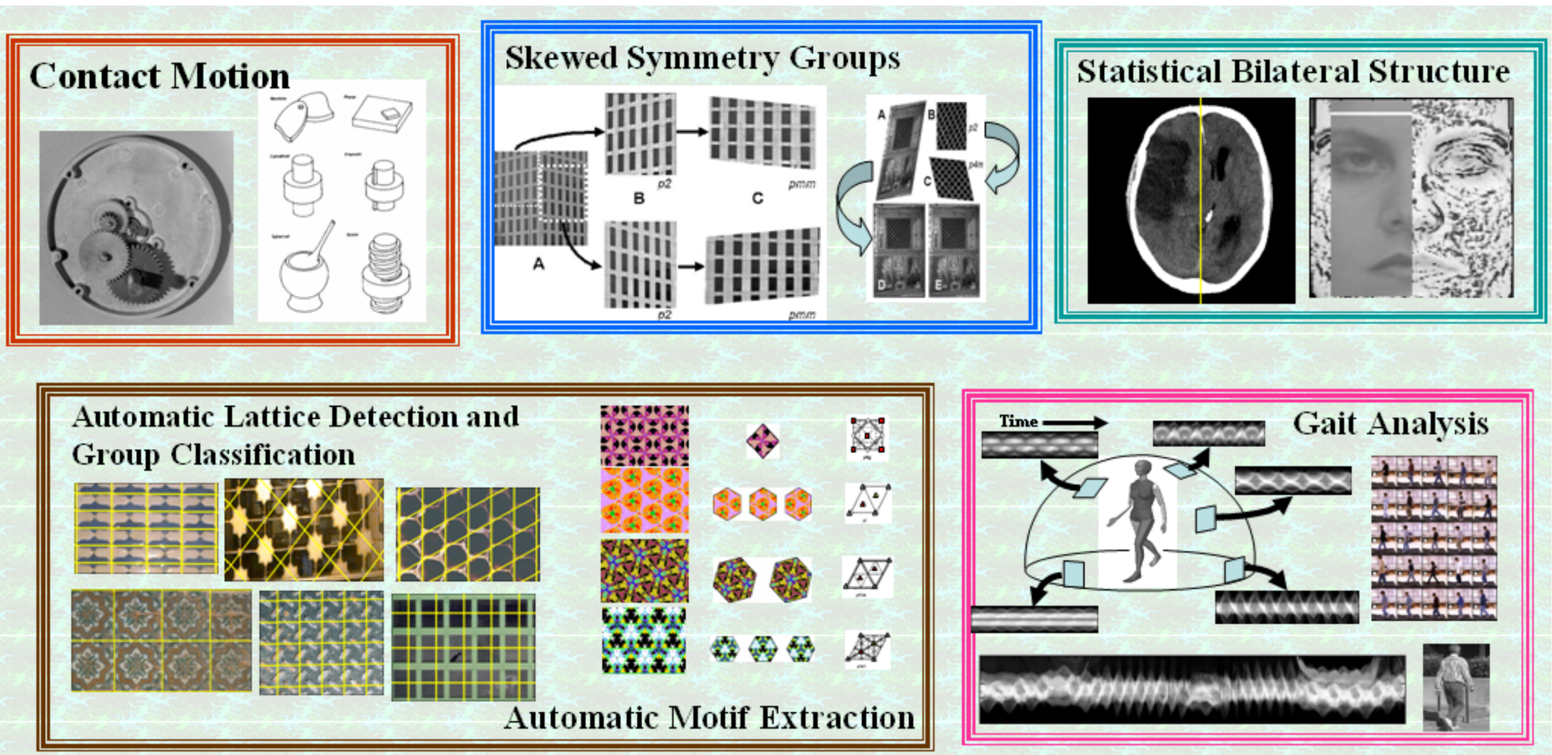
Functions of Bayesian Reasoning

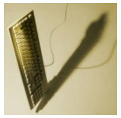
- Bayesian learning techniques are prominent in fields such as computer vision, bioinformatics, and natural language processing (including, but not limited to spam filtering)

Computer Vision

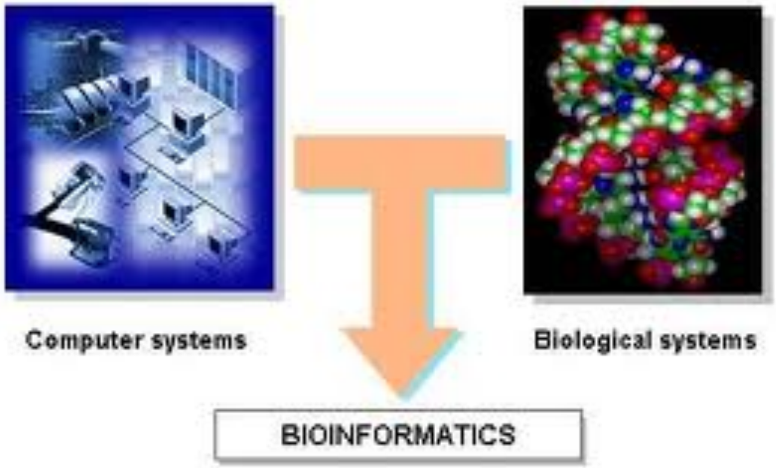


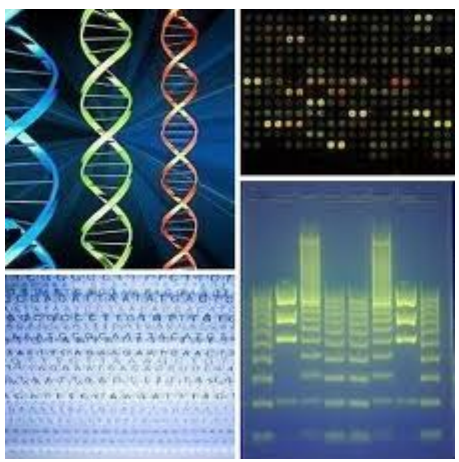


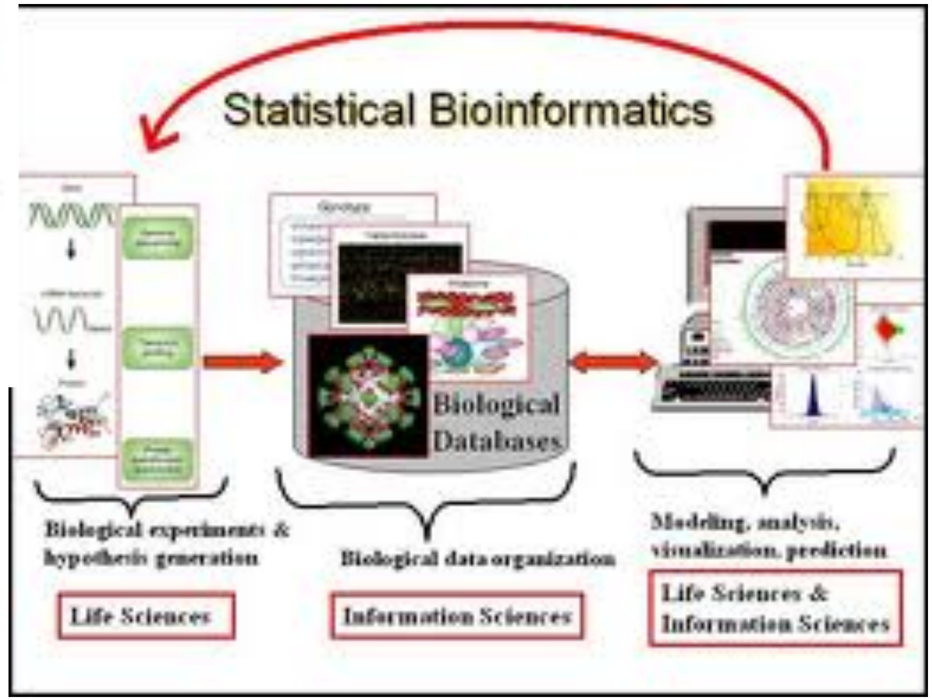




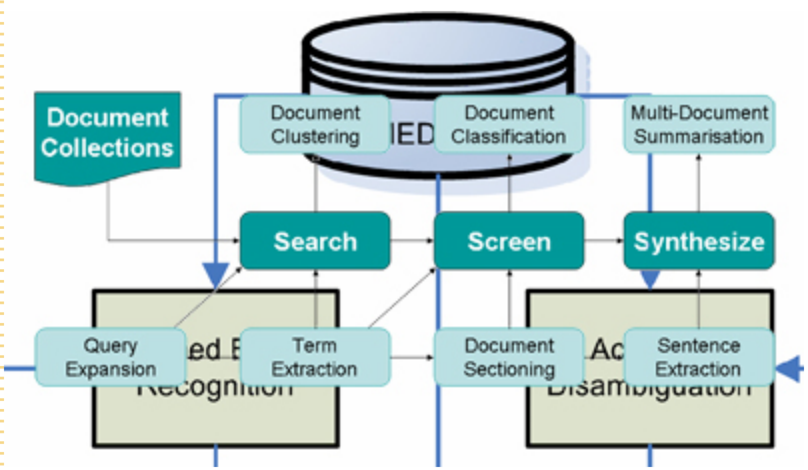
Statistical Bioinformatics

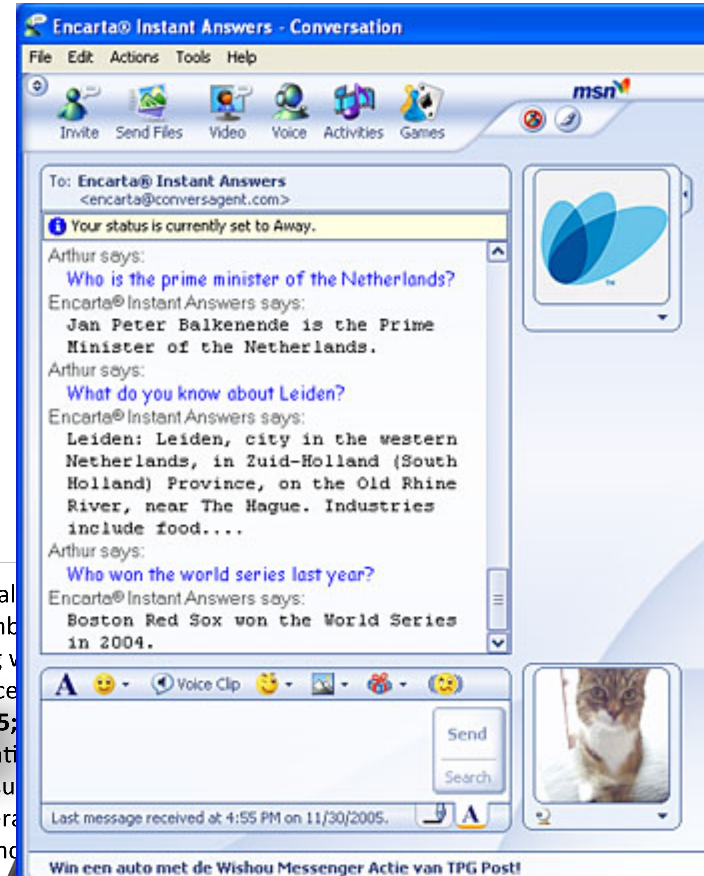






Natural Language Processing





BlackBoard SafeAssignment

[5; 97%] Knowledge Representation Knowledge representation is an area in artificial intelligence that is concerned with how to formally "think", that is, how to use a symb system to represent "a domain of discourse" - that which can be talked about, along w functi nce (form **[5; 100%]** Generally speaking, some kind of logic is used both to supply a formal semanti how reasoning functions apply to symbols in the domain of discourse, as well as to su (depending on the particulars of the logic), operators such as quantifiers, modal opera etc. **[5; 100%]** that, along with an interpretation theory, give meaning to the sentenc the logic.

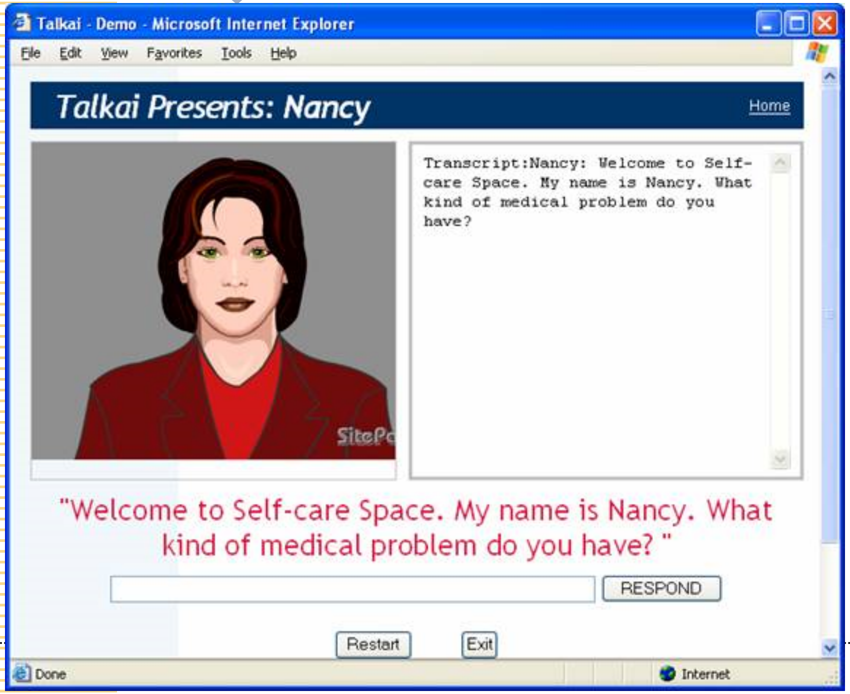
[5; 93%] The key problem is to find a knowledge representation (and a supporting reasoning system) that can make the inferences your application needs in time, tha within the resource constraints appropriate to the problem at hand. **[5; 100%]** This tension between the kinds of inferences an application "needs" and what counts as time" along with the cost to generate the representation itself makes knowledge representation engineering interesting.

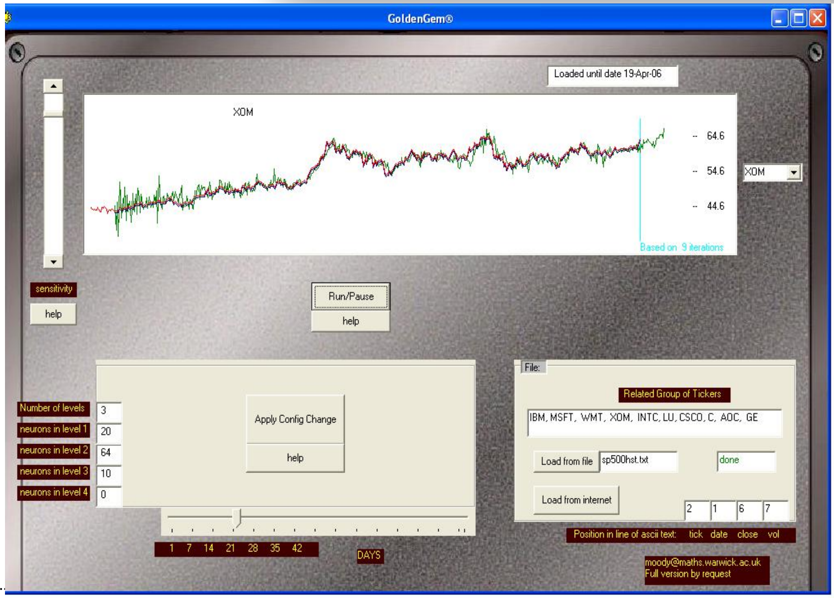
[5; 94%] There are representation techniques such as frames, scripts, objects, rules, tagging, and semantic networks which have originated from theories of human information processing. **[5; 86%]** Since knowledge is used to achieve intelligent behavior, the fundamental goal of knowledge representation is to represent knowledge in a manner as to facilitate inference from knowledge.

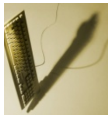
Other Applications

Card games that has to decide actions to take when playing with you.

Expert systems

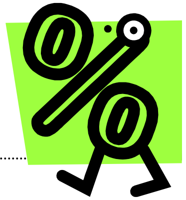


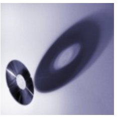


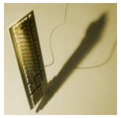


Basic probability theory

- The **probability** of an event is the proportion of cases in which the event occurs. Probability can also be defined as a ***scientific measure of chance***.







Probability Theory

- Expressed as a numerical index
- range

0 ---------------------------------- 1

absolute impossibility absolute certainty

- Example: Is it raining?

0 0.5 1

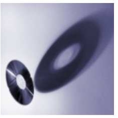
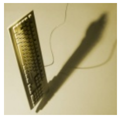
Not rain I don't know Rain

Probability

- each event has at least two possible outcomes: favorable outcome (success) and unfavorable outcome (failure).

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

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

Probability

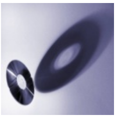
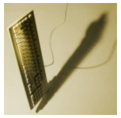
- s : the number of times success can occur
- f : the number of times failure can occur

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

and

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

$$p + q = 1$$

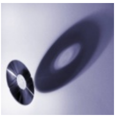
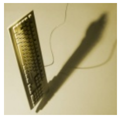




Example – tossing a coin

- P(head) = P(tail)
- i.e. in a single throw, the probability of getting a head (or a tail) is 0.5.

Conditional probability

- Consider 2 events: *A* and *B*
  - *A = Road is wet, B = raining*
- Assumption: *A* and *B* are not mutually exclusive
- A occur conditionally on the occurrence of B.
  - *Road is wet because it was raining*
- The probability that event *A* will occur if event *B* occurs is called the **conditional probability**.

Conditional Probability

- $p(A|B)$ = "Conditional probability of event A occurring given that event B has occurred".

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

where,

the probability that *A* and *B* can occur = $p(A \cap B)$

the number times B can occur = $p(B)$

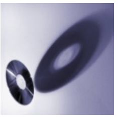
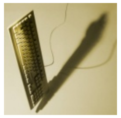
Conditional Probability (cont)

- Therefore

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

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

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

Conditional Probability (cont)

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

Hence,

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

Substituting the last equation into the equation

Conditional Probability (cont)

the **Bayesian rule / Baye's Theorem**:

$$\boxed{p(A|B) = \frac{p(B|A) \times p(A)}{p(B)}}$$

Bayesian rule

$$p\left(A \middle| B\right) = \frac{p\left(B \middle| A\right) \times p\left(A\right)}{p\left(B\right)}$$

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.

Types of Probabilities

prior probabilities: $p(H)$ and $p(\sim H)$

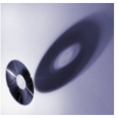
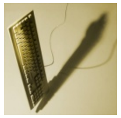
- Probability of hypothesis H

conditional probabilities : $p(E|H)$, $p(E|\sim H)$

- Probability of evidence E if hypothesis *H* is true

posterior probability : $p(H|E)$

- Probability of hypothesis *H* upon observing evidence *E*

Bayesian Rule

- expressed in terms of hypotheses (H) and evidence (E)

$$P(H \mid E) = \frac{P(E \mid H) * P(H)}{P(E)}$$

where:

$p(H|E)$ is the posterior probability that hypothesis H occurs given that evidence E has occurred;
$p(E|H)$ is the conditional probability of evidence E occurring given that hypothesis H has occurred;
$p(H)$ is the prior probability of hypothesis H occurring;
$p(E)$ is the probability of evidence E occurring.

Exercise

- IF it was raining yesterday night
  THEN the backyard is wet in the next morning (P=0.7)
- It rains 200 days in a year.
- There is 60% chance that the backyard will be wet
- Question: If the backyard is wet, what is the chance that yesterday night rained? P(rain|wet) = ?

$$
\begin{aligned}
P(rain|wet) &= P(wet|rain)*P(rain) / P(wet) \\
&= 0.7*(200/365) / 0.6 \\
&= 0.64
\end{aligned}
$$

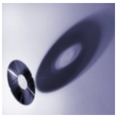
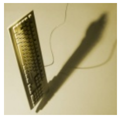
Conclusion: 64% it was raining yesterday night

IF P(B) is not known

- Another form of Bayes Theorem:

$$P(A/B) = \frac{P(B/A)*P(A)}{P(B/A)*P(A) + P(B/not(A))*P(not(A))}$$

- P(A/B) = probability of Event A occurring , given that B has already occurred.
- P(A) = probability of Event A occurring (prior probability), without evidence.
- P(B/A) = additional evidence of B occurring, given A.
- P(not A) = A is not going to occur.
- P(A) + P(not A) = 1

If the occurrence of event *A* depends on only two mutually exclusive events, *B* and NOT *B* , we obtain:

$$p(A) = p(A|B) \times p(B) + p(A|\neg B) \times p(\neg B)$$

where ¬ is the logical function NOT.

Similarly,

$$\boxed{p(A|B) = \frac{p(B|A) \times p(A)}{p(B)}}$$

The Bayesian rule

expressed in terms of hypotheses (H) and evidence (E)

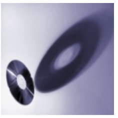
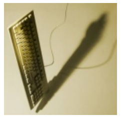
$$p(H|E) = \frac{p(E|H) \times p(H)}{p(E|H) \times p(H) + p(E|\neg H) \times p(\neg H)}$$

Example of Problem

- P(patient-has-measles | patient-has-spots)
- Given the data:

  50%of people have spots because of measles. There is a 1% chance of someone having measles in the world (given no evidence for or against).
- That is:

  P(spots | measles) = 0.5
  P(measles) = 0.01

Solution

$$P(A/B) = \frac{P(B/A)*P(A)}{P(B/A)*P(A) + P(B/not(A))*P(not(A))}$$

- P(measles | spots)=

$$\frac{P(spots \mid measles) * P(measles)}{P(spots|measles)*P(measles) + P(spots|not\ measles)*P(not\ measles)}$$

P(spots|measles) = 0.5 ==> P(spots|not measles) = 1-0.5 = 0.5

P(measles) = 0.01 ==> P(not measles) = 1-0.01 = 0.99

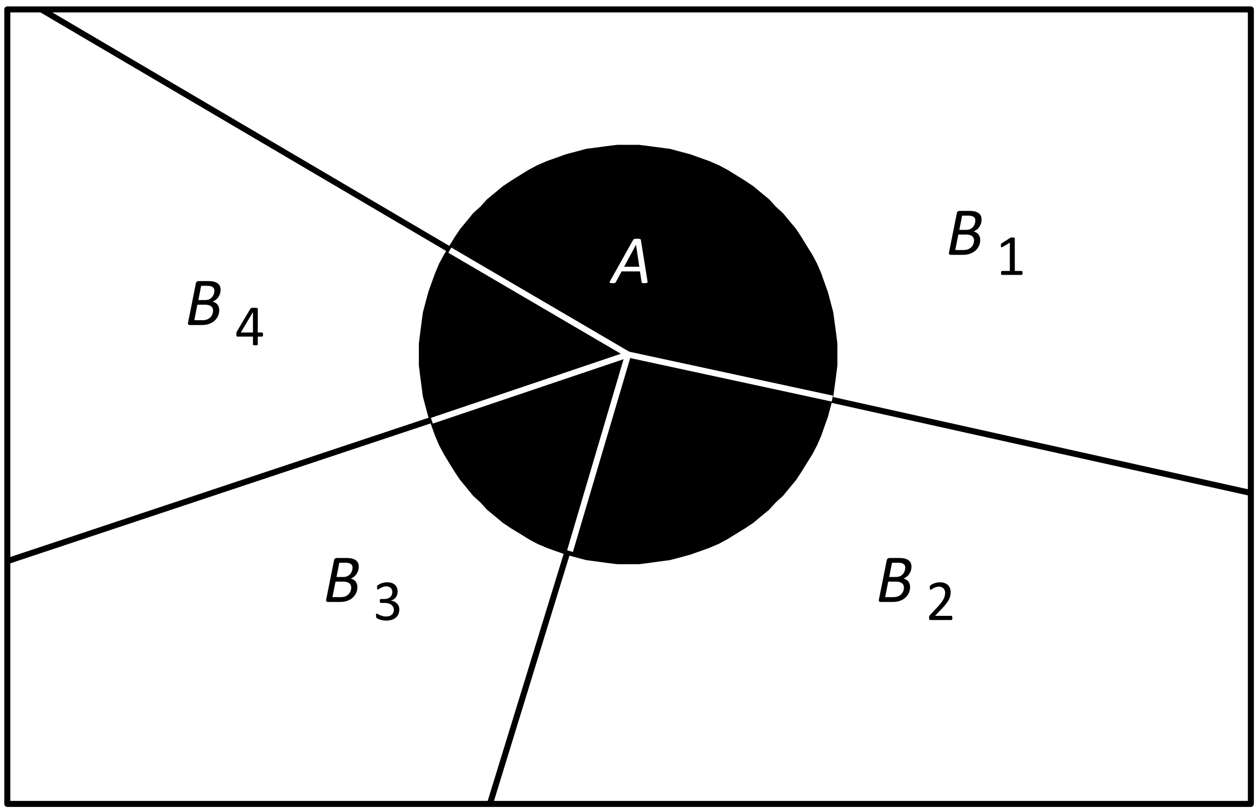
Therefore:

P(measles|spots) = 0.5*0.01 /((0.5*0.01) + (0.5*0.99))= 0.01

→ Only 1% chance that the patient has measles if red spots found

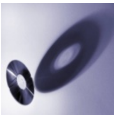
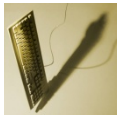
The joint probability

$$\sum_{i=1}^{n} p\left(A \cap B_i\right) = \sum_{i=1}^{n} p\left(A \middle| B_i\right) \times p\left(B_i\right)$$



Multiple Hypotheses and Evidences

- We can take into account both multiple hypotheses H_1, H_2,..., H_m and multiple evidences E_1, E_2 ,..., E_n.
- The hypotheses as well as the evidences must be mutually exclusive and exhaustive.

Single E and Multiple H

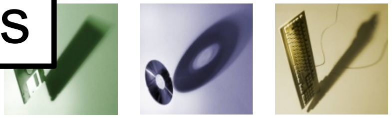
- Single evidence *E* and multiple hypotheses follow:

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

evidence

- E.g.
  – a patient is observed to have **red spots**.
  – He can be infected by **dengue, chicken pox, measles**, etc.

hypotheses



Exercise

- In ABC College, of all their students, 25% obtained good results, 50% are average, and 25% are bad. Suppose for any examination, a student who obtained good results has a 5% chance of failing the examination, an average student has 15% chance failing an exam, and a student had bad results has a 25% chance that will fail.
- If you failed an examination in the past year, what is the probability that you can obtain good result for the coming exam?

Answer

$p(good) = 0.25$

$p(average) = 0.50$

$p(bad) = 0.25$

$p(fail|good) = 0.05$

$p(fail|average) = 0.15$

$p(fail|bad) = 0.25$

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

$$p(good|fail) = \frac{p(fail|good) * p(good)}{p(fail|good).p(good) + p(fail|average).p(average) + p(fail|bad).p(bad)}$$

$$= (0.05)(0.25) \,/\, (\,(0.05)(0.25) + (0.15)(0.50) + (0.25)(0.25)\,)$$

$$= 0.083$$

Multiple E and Multiple H

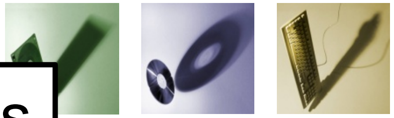
- Multiple evidences and multiple hypotheses follow:

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

evidences

- E.g.
  - a patient is observed to have **red spots, fever and flu-like symptoms**.
  - He can be infected by **dengue, chicken pox, measles**, etc.

hypotheses



Multiple E and Multiple H

- Multiple evidences and multiple hypotheses follow:

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

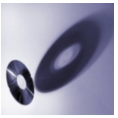
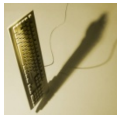
This requires to obtain the conditional probabilities of all possible combinations of evidences for all hypotheses

Which is very difficult

Multiple E and Multiple H

- Therefore **conditional independence** among different evidences assumed. Thus:

$$p\left(H_i \middle| E_1\, E_2 \ldots E_n\right) = \frac{p\left(E_1 \middle| H_i\right) \times p\left(E_2 \middle| H_i\right) \times \ldots \times p\left(E_n \middle| H_i\right) \times p\left(H_i\right)}{\sum_{k=1}^{m} p\left(E_1 \middle| H_k\right) \times p\left(E_2 \middle| H_k\right) \times \ldots \times p\left(E_n \middle| H_k\right) \times p\left(H_k\right)}$$

Multiple E and Multiple H

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

Assume that we first observe evidence E_3. The expert system computes the posterior probabilities for all hypotheses as

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

Thus,

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

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

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

Suppose now that we observe evidence E_1.

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

Hence,

$$p\left(H_1 \middle| E_1 E_3\right) = \frac{0.3 \cdot 0.6 \cdot 0.40}{0.3 \cdot 0.6 \cdot 0.40 + 0.8 \cdot 0.7 \cdot 0.35 + 0.5 \cdot 0.9 \cdot 0.25} = 0.19$$

$$p\left(H_2 \middle| E_1 E_3\right) = \frac{0.8 \cdot 0.7 \cdot 0.35}{0.3 \cdot 0.6 \cdot 0.40 + 0.8 \cdot 0.7 \cdot 0.35 + 0.5 \cdot 0.9 \cdot 0.25} = 0.52$$

$$p\left(H_3 \middle| E_1 E_3\right) = \frac{0.5 \cdot 0.9 \cdot 0.25}{0.3 \cdot 0.6 \cdot 0.40 + 0.8 \cdot 0.7 \cdot 0.35 + 0.5 \cdot 0.9 \cdot 0.25} = 0.29$$

Hypothesis H_2 has now become the most likely one.

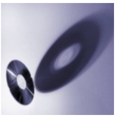
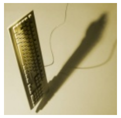
After observing evidence E_2, :

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

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

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

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

Exercise

- Assume that the chance of a person sneezing caused by flu virus infection is 0.3 and the chance of a person sneezing caused by allergy rhinitis is 0.95. According to the medical reports of all students in a school, 10% of the students had allergic rhinitis and 20% of the students had flu in the past 6 months.

Assume that you observed a student in the school sneezing in the class. Demonstrate the working using the Bayesian Theorem, generate the conclusion with its respective conditional probability.

Remarks: Bayes' theorem for multiple hypotheses:

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

Exercise

- Assume that the chance of a person sneezing caused by flu virus infection is 0.3 and the chance of a person sneezing caused by allergy rhinitis is 0.95. According to the medical reports of all students in a school, 10% of the students had allergic rhinitis and 20% of the students had flu in the past 6 months.

Assume that you observed a student in the school sneezing in the class. Demonstrate the working using the Bayesian Theorem, generate the conclusion with its respective conditional probability.

Remarks: Bayes' theorem for multiple hypotheses:

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

Next

- Certainty Factor

Certainty Factor (CF)

a number in the range -1 to 1 (usually)

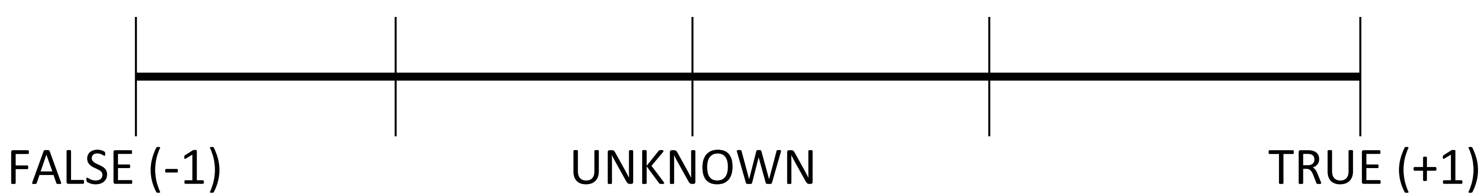
negative indicates disbelief.

- -1 indicates maximum disbelief

Positive indicates belief

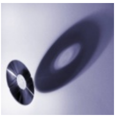
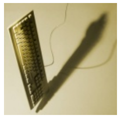
- 1 indicates maximum belief

0 indicates ‘don't know’



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

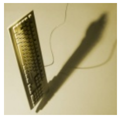
- An expert system can propagate CFs via rules to give CFs to conclusions. For this, rules too may have CFs associated with them. E.g.

  if (B or C) then A (CF 0.7)

- This says that if we know **B (cf = 1)** or **C (cf=1)** to be true then this knowledge will give us a degree of belief that **A** is true.

But what if **B** and **C** themselves are subject to uncertainty?





Formula

- We must provide rules which allow CFs of logical combinations to be calculated.
- Rules:
  - ***CF(not(A)) = −CF(A)***
  - ***CF(A and B) = min(CF(A),CF(B))***
  - ***CF(A or B) = max(CF(A),CF(B))***

Example 1

- CF(not rain) = 0.8,
- What is the CF for CF(rain)??

<u>Solution:</u>

Example 2

- IF you did exercise **and** did assignment, THEN you will pass exam.
- Given:

  CF(did exercise) = 0.8

  CF(did assignment)=0.9

<u>Solution:</u>

Example 3

IF you did exercise **OR** did assignment,
THEN you will pass exam. **(CF 0.9)**

Given:

CF(did exercise) = 0.8

CF(did assignment)=0.7

<u>Solution:</u>

Example 4

- Here is a set of rules:

  if (B or C) then A (CF 0.6)

  if (D and E and F) then B (CF 1.0)

  if (G or H) then C (CF 0.75)

- Suppose that we know CFs for D, E, F, G and H as follows:

  CF(D) = 0.8 CF(E) = 0.5

  CF(F) = 0.9 CF(G) = 0.2

  CF(H) = 0.8

- Find the CF for A.

Solution

CF(D) = 0.8

CF(E) = 0.5

CF(F) = 0.9

CF(G) = 0.2

CF(H) = 0.8

Certainty Factor(Different evidence/rule with the same conclusion)

- It may happen that a conclusion is supported by more than one rule. For example, in addition to the rules above, we might have another rule for A:
-     if (P or Q) and R then A (CF 0.8)
- This could lead to a quite independent calculation of a CF for A. How should this be combined with the other one to obtain an overall CF?

Formula(Combination of 2 Rules)

- Suppose that we can find certainty factors *CF1* and *CF2* for a conclusion by application of two independent rules.
- We can then calculate an overall CF by applying one of three formulas:

$$CF1 + CF2 - CF1*CF2 \quad \text{if both are positive,}$$

$$CF1 + CF2 + CF1*CF2 \quad \text{if both are negative,}$$

$$\frac{CF1 + CF2}{1 - min(|CF1|,|CF2|)} \quad \text{otherwise}$$

$$
cf(cf_1, cf_2) = \begin{cases} cf_1 + cf_2 x\,(1 - cf_1) & \text{both } cf > 0 \\[2ex] \dfrac{cf + cf}{1 - \min[\,|cf_1|\,], [\,|cf_2|\,]} & \text{one } cf < 0 \\[2ex] cf_1 + cf_2 x\,(1 + cf_1) & \text{both } cf < 0 \end{cases}
$$

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

Example

- R1: if you studied hard, then you pass exam.
  CF = 0.36 => A1
- R2: if you did tutorials, then you pass exam.
  CF = 0.75 => A2
- R3: if you pass assignment, then you pass exam.
  CF = 0.5 => A3
- Assume that ‘you studied hard’, ‘you did tutorials’ and ‘you pass assignment’ have cf. {1.0}
- What is CF(pass exam)??

Solution:

Exercise

R1: IF nose is stuffy [P1.1] **AND** stuffy nose resolves within week [P1.2] THEN symptom of cold [H1]	**R3:** IF sneezing present [P3.1] **AND** headaches not present [P3.2] **AND** sore throat presents [P3.3] THEN symptom of cold (cf(H3) = 0.9)
R2: IF cough present [P2.1] **AND** cough is productive [P2.2] THEN symptom of cold [H2]	**R4:** IF fever not present [P4.1] **OR** fever is mild [P4.2] THEN symptom of cold (cf(H4) = 0.8)

- His nose is stuffy CF = 0
- Stuffy nose resolves within week CF = 0
- Cough present CF = 0.8
- Cough is productive CF = -0.4
- Sneezing present CF= 1.0

- **Headaches present CF = 1.0**
- Sore throat present CF = 0.8
- **Fever present CF = 1.0**
- Fever is mild CF = 0.4

identify the certainty factor that the user is getting a cold.

Solution

- CF(R1)=
- CF(R2)=
- CF(R3)=
- CR(R4)=

Solution (Cont)

- Now we combine rule by rule:
- We can ignore R1 since it is 0.
- CF(R2, R3) =
- CF(R2,R3,R4) =
- Conclusion: The user probably NOT getting cold since CF(cold) **= -0.91**
- *Remark: in this case, the user may get flu*

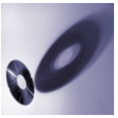
Advantages

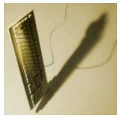
Easy to compute

- can be used to easily propagate uncertainty through the system.

It has been found to be remarkably resilient

- Systems which use them seem to work reasonably well in practice, in comparison with human experts.





Drawbacks

Assumption of independent probabilities

- The certainty factor of two rules in an inference chain is calculated as independent probabilities.

CF is not being rigorously founded

- The use of CFs has been criticized (with some justification) for not being rigorously founded.