Certainty Factor Models (CFM)

Material used

- Frans Voorbraak: Certainty Factors (in the reader, lecture)

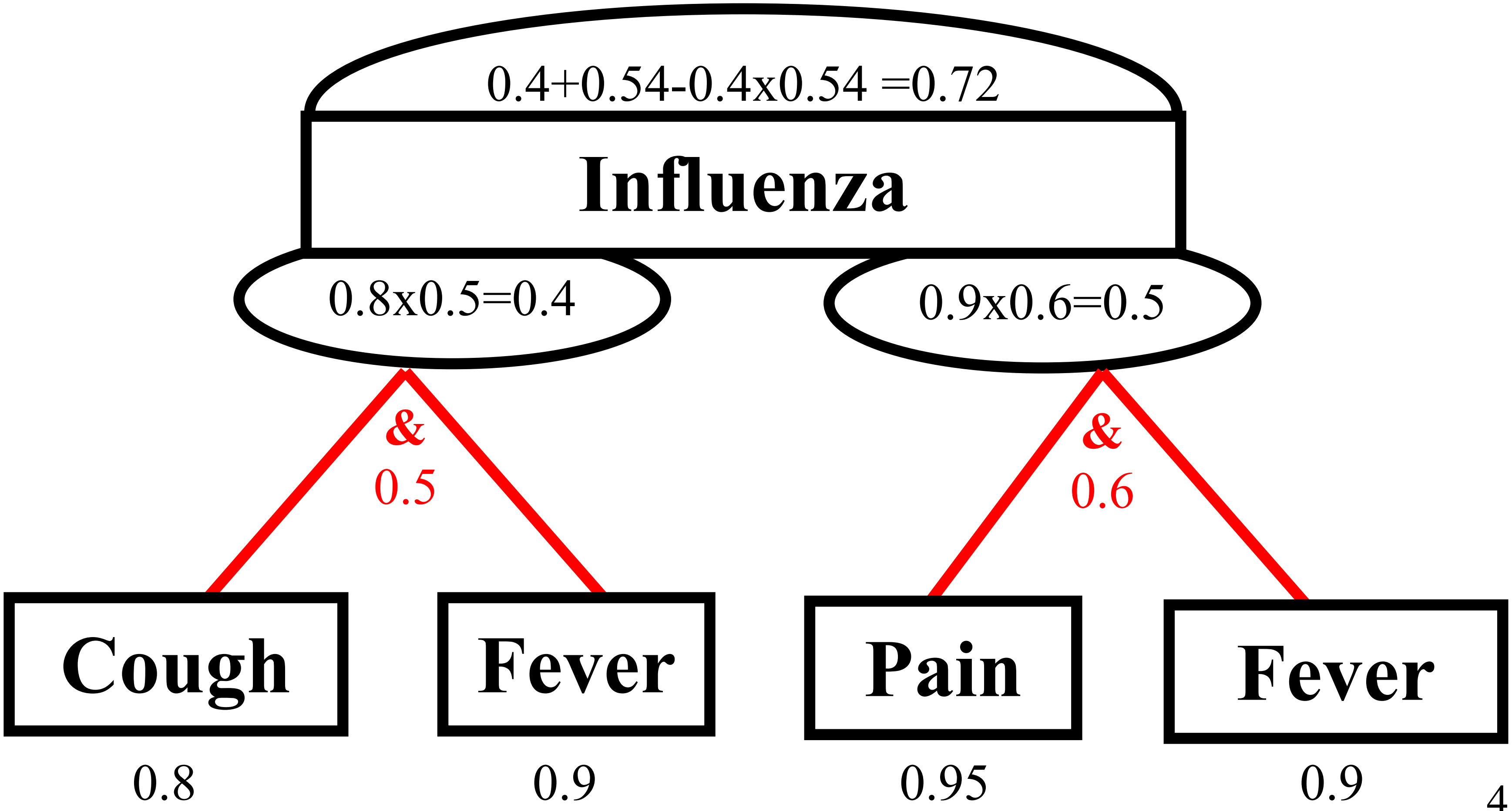
Methods for Dealing with Uncertainty

- **1. Numeric Methods**
  - *Bayesian Model*
  - *Certainty Factor Model*
  - *Dempster - Shafer Theory*
  - *Possibility Theory*
- **2. Fuzzy Logic**
- **3. Non - Monotonic Logic**

Certainty Factors

- A Certainty Factor is a numerical value that expresses the extent to which, based on a given set of evidence, we should accept a given conclusion. A Certainty Factor or CF with a value of 1 indicates total belief, whereas a CF with a value of -1 indicates total disbelief.
- In a system that uses CFs, the rules must be so structured that any given rule either adds to belief in a given conclusion or adds to disbelief.

Elementary illustration



General characteristics of the CFM

- Developed for the rule-based expert system MYCIN
- Very simple model specifically aimed at use with rule-based systems
- Much used in practice, since rule-based systems are still popular
- To some degree successfull
- Heavily criticised because of lack of theoretical justification
- Now facing heavy competition from Bayesian Networks (as is the rule-based formalism)

Certainty Theory

- Representing uncertain evidence (facts)
- Representing uncertain rules
- Combining evidence from multiple sources

  eg. IF A and B THEN X [CF= 0.8]

  IF C THEN X [CF=0.7]

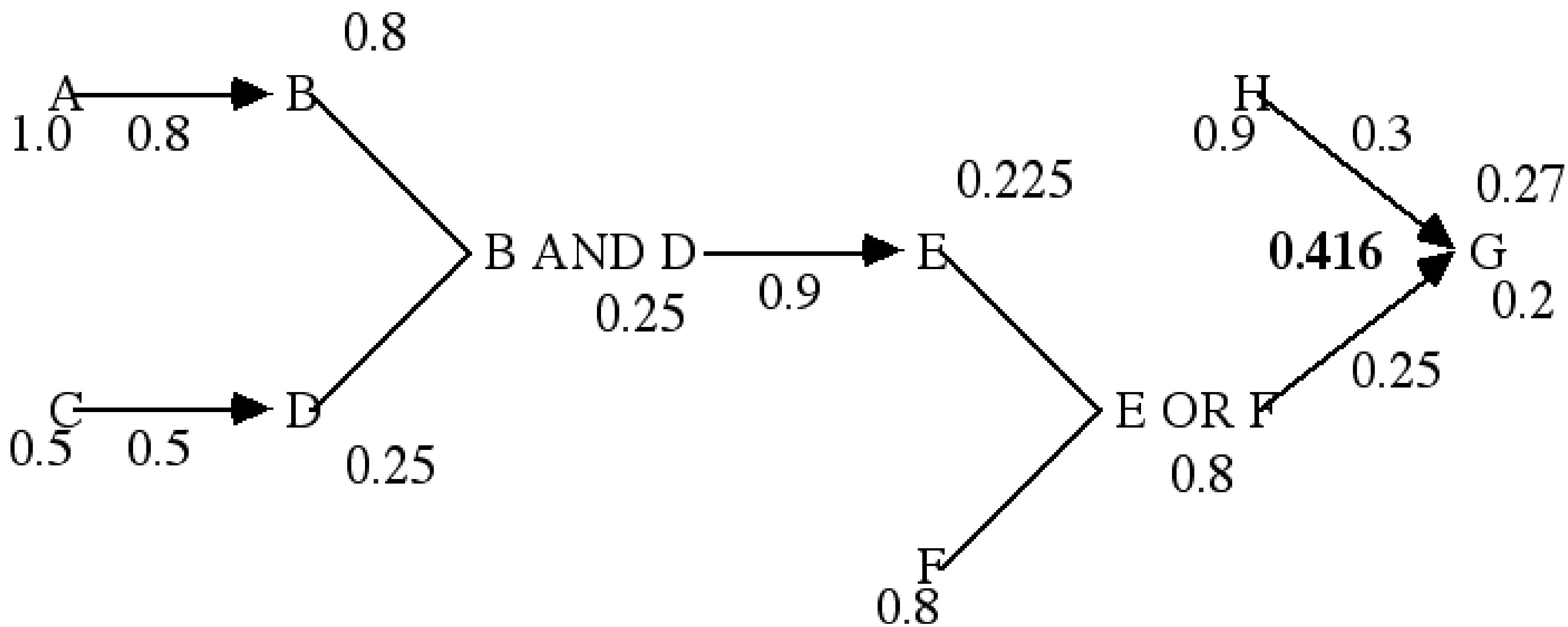
  What is the certainty of X? (given the CFs of A,B,C)

- -1<= CF <= 1
- Users provide CFs for specific facts; Experts provide CFs for rules
- Explanation of CFs in terms of Probabilities?

Some notations

▪ CF(A)	CF of a uncertain evidence
▪ CF(IF A THEN B) also CF(A → B)	CF of a rule
▪ CF(X|R)	CF of X given the rules in the set R (+ required facts) X is a non-initial fact

Inference network



ALL branches leading to G are needed to determine its CF!

Calculations

CF(B) = CF(A)*CF(IF A THEN B) = 1*.8 = .8

CF(D) = CF(C)*CF(IF C THEN D) = .5*.5 = .25

CF(B&D) = min(CF(B),CF(D)) = min(.8,.25) = .25

CF(E) = CF(B&D)*CF(IF B&D THEN E) = .25*.9 = .225

CF(E OR F) = max(CF(E),CF(F)) = max(.225,.8) = .8

CF(G | {IF E OR F THEN G}) =

CF(E OR F)*CF(IF E OR F THEN G) = .8*.25 = .2

CF(G | {IF H THEN G}) =

CF(H)*CF(IF H THEN G) = .9*.3 = .27

CF(G) = CF(G | {IF E OR F THEN G, IF H THEN G}) = .2 + .27 - .2*.27 = .416

Propagation rules

$CF(A \wedge B) = \min(CF(A), CF(B))$

$CF(A \vee B) = \max(CF(A), CF(B))$

$CF(B \mid \{IF\ A\ THEN\ B\}) = CF(A\ THEN\ B) * \max(0, CF(A))$

If $CF(B \mid R) = x$, $CF(B \mid S) = y$, and R and S have no rules in common and are both non-empty, then

$$CF(B \mid R \cup S) = \begin{cases} x + y - xy & \text{if } x > 0, y > 0 \\ x + y + xy & \text{if } x < 0, y < 0 \\ \dfrac{x + y}{1 - \min\{|x|, |y|\}} & \text{otherwise} \end{cases}$$

Motivation of the combination rules

- Antecedent pooling: from fuzzy logic
- Serial combination: the factor max(0,CF(A)) avoids that evidence against A can have an effect on the certainty of B through A➔B
- The parallel combination function *f* is an improved version of the following combination function which was originally used:

$$g(x,y) = \begin{cases} x + y - xy & \text{if } x, y > 0 \\ x + y + xy & \text{if } x, y < 0 \\ x + y & \text{otherwise,} \end{cases}$$

(note that it is undefined for x = 1, y= -1, and for x = -1, y=1)

Illustrating example

R1: IF weatherman says it will rain
THEN it will rain CF 0.8

R2: IF farmer says it will rain
THEN it will rain CF 0.8

Case (a): Weatherman and farmer are certain in rain

$CF(E_1) = CF(E_2) = 1.0$

$CF(H, E_1) = CF(E_1) * CF(Rule_1) = 1.0*0.8 = 0.8$

$CF(H, E_2) = CF(E_2) * CF(Rule_2) = 1.0*0.8 = 0.8$

$CF_{combine}(CF_1, CF_2) = CF_1 + CF_2(1 - CF_1) = 0.8+0.8(1-0.8) = 0.96$

CF of a hypothesis which is supported by more than one rule, can be incrementally increased by supporting evidence from both rules.

Illustrating example, continued

Case (b): Weatherman certain in rain, farmer almost certain in no rain

$CF(E_1) = 1$, $CF(E_2) = -0.99$

$CF_1 = 0.8$, $CF_2 = -0.792$

$CF_{combine}(CF_1, CF_2) = (0.8 + (-0.792)) / (1 - \min(0.8, 0.792)\} = 0.04$

Case (c): Incremental decrease in certainty from more than one source of disconfirming evidence

$CF_{combine}(CF_1, CF_2, CF_3, \ldots) = 0.999 = CF_{old}$

Single piece of disconfirming evidence $CF_{new} = -0.8$

$Cf_{combine} = (0.999 - 0.8) / (1 - 0.8) = 0.995$

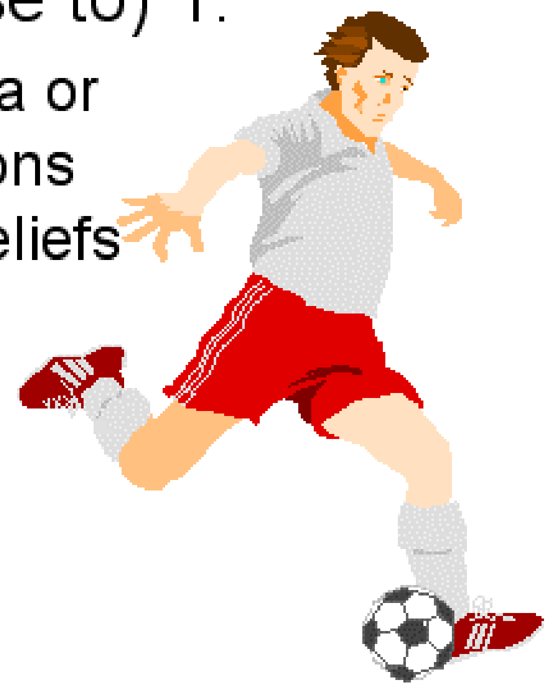
Single piece of disconfirming evidence does not have a major impact on many pieces of confirming evidence.

Problems with CFs

CF(AvB) is (close to) 1 if and only if

CF(A) is (close to) 1 or CF(B) is (close to) 1.

- Being (almost) certain that FC Barcelona or Manchester United will win the Champions League does not imply having strong beliefs about which of the two teams will win!



Problems with CFs, continued

Assume CF(A) = 0.4, CF(B1) = 0.3, CF(B2) = 0.6

Rule Base 1:

{IF A∧(B1∨B2) THEN C (1.0)}

⇨ CF(C) = min(0.4,max(0.3,0.6)) = 0.4

Rule Base 2:

{IF A∧B1 THEN C (1.0), IF A∧B2 THEN C (1.0)}

⇨ CF(C) = 0.3 + 0.4 - 0.3*0.4 = 0.58

Suppose a domain expert believes with 70 % certainty that *gram positive cocci growing in chains* are *streptococci*.

What is an appropriate measure of belief to represent this knowledge?

$MB(E \rightarrow H) = P(H|E)$??

Problem: $MB(E \rightarrow H) = 0.7$, then $MB(E \rightarrow \sim H) = 0.3$

Foundation, continued

The Measure of Belief (MB) is a number that reflects the measure of **increased** belief in a hypothesis *H* based on evidence *E*. (Shortliffe & Buchanan)

$$\mathrm{MB}(E \rightarrow H) = \begin{cases} 1 & \text{if } P(H) = 1 \\ \dfrac{P(H \mid E) - P(H)}{1 - P(H)} & \text{if } P(H|E) \\ 0 & \text{otherwise} \end{cases}$$

$$\mathrm{CF}(E \rightarrow H) = \mathrm{MB}(E \rightarrow H) - \mathrm{MB}(E \rightarrow \sim H)$$

Foundation, continued

Obvious consequence:

$$CF_P[E \longrightarrow H] = \begin{cases} 1 & if\ P(H) = 1 \\ \frac{P(H|E)-P(H)}{1-P(H)} & if\ P(H|E) \geq P(H) \notin \{0,1\} \\ \frac{P(H|E)-P(H)}{P(H)} & if\ P(H|E) \leq P(H) \notin \{0,1\} \\ -1 & if\ P(H) = 0. \end{cases}$$

So far, CF_P is only defined for rules. In order to apply it to facts we can define $CF_P(A) = CF_P(e \rightarrow A)$ where e is the external evidence available to the user.

Foundation of propagation?

It can be shown that, given the above probabilistic interpretation of CF's , the propagation rules of the CT model are only valid under trivializing assumptions.

Perhaps they are reasonable approximations in some cases. The following example shows that this is a rather weak line of defence.

Counterexample

Let $\Omega = \{1,2,3,4,5,6\}$. Then define for every $X \subseteq \Omega$, $P(X) = |X|/6$.

Assume the user's evidence E is represented by {1,2}. Further, assume A is represented by {1} and B is represented by {2}.

Then $CF_P(A) = CF_P(E \rightarrow A) = (1/2 - 1/6)/(1-1/6) = 0.4$. Similarly, $CF_P(B) = 0.4$.

If $CF_P(A) = CF_P(B) = 0.4$, then according to the propagation rules $CF(A \wedge B) = CF(A \vee B) = 0.4$.

However, $CF_P(A \wedge B) = -1$ and $CF_P(A \vee B) = 1$.

Evaluation

+ Nicely fits the rule-based paradigm, no essential change in representation formalism necessary.
+ Easy to implement.
+ Some practical successes.
+ Propagation rules at first sight not unreasonable.

– Interpretation, meaning of CF unclear.
– No decision theory associated with CF.
– Propagation rules cannot be justified.
– Successes of CF can be explained to be largely independent of CF model.