Master of Technology in Knowledge Engineering

KE4102: Intelligent Systems & Techniques for Business Analytics: Uncertainty Handling in RBS

Charles Pang
Institute of Systems Science
National University of Singapore
E-mail: charlespang@nus.edu.sg

Contents

- Sources of Uncertainty
- Managing uncertainty in rule-based systems
- Basic probability
- Bayes Theorem
- Certainty Factors
- Conceptual design of RBS with CF

Uncertainty in human reasoning

- Uncertainty is defined as <u>the lack of exact knowledge</u> that would enable us to reach a perfectly reliable conclusion
- The real world is imperfect and knowledge is often incomplete, inconsistent, uncertain ... but humans have the ability to make “correct” decisions
- Can Artificial Intelligence Systems be made to deal with these uncertainties and help us make good decisions?

Sources of uncertainty

- Theory of Domain maybe vague:
  – dealing with phenomena that are imperfectly understood
- Captured data may be incomplete:
  – erroneous, missing data
  – considerations of cost and risk deter their acquisition (e.g. expensive medical test/surgery)
- Being human
  – concepts are not precisely defined
  – the use of rules-of-thumb, best guess, hunches

How do we cope with uncertainty?

- Use of heuristics
  - Might help to mitigate the uncertainty
- Approximate the data
  - obtain data by proxy means
- Mining the data for some "certainty"
  - Find patterns and associations that can be useful
- Quantify the meaning of vague words
  - In 1944, Ray Simpson asked 355 high school and college students to place 20 terms like "often" on a scale between 1 and 100. In 1968, Milton Hakel repeated this experiment

Quantification of imprecise terms

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

Managing Uncertainty in RBS

There are 3 main considerations when implementing an uncertainty scheme:

1. How to **represent** uncertain Knowledge & data
2. How to **combine** two or more pieces of uncertain knowledge & data
3. How to draw **inference** using uncertain knowledge & data

Methods for Managing Uncertainties

There are broadly 3 methods for managing uncertainties in Rule based reasoning:-

1. Symbolic method
   – True/False/Neither True nor False
2. Statistical or Mathematical methods
   – **Bayes Theorem, Certainty Factors**
3. Fuzzy Logic Method
   – Fuzzy Set Theory

NUS National University of Singapore | ISS INSTITUTE OF SYSTEMS SCIENCE
© 2018 NUS. All Rights Reserved

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
- Many aspects of uncertainty can be represented as probabilities
- For example, if event A occurs N_a times out of a total of N occasions, then the probability of event A is

$$p(A) = \frac{N_a}{N}$$

E.g. $$\textbf{\textit{p}}\,(\textit{being late for class}) = \frac{\textit{number of times late for class}}{\textit{number of classes}}$$

Deriving probability values

- Sometimes we use statistics/mathematics; e.g. probability of getting two 6's with the throw of 2 dices (i.e. p(12)= 1/36)
- Sometimes we estimate the probabilities using empirical means; e.g. conduct experiments multiple times (throw dice 1000x or use surveys such as the Simpson & Hakel experiments)
- Sometimes we simply use our experience/best guess; e.g. attaching weights to chunk of knowledge (the chances of engine failure due to over-heating). There is no systematic method to obtain this weight – it's just a "hunch"

Conditional Probability

- Conditional probability is the most widely used measure for solving many types of problems
- **Conditional probability** is defined as the probability of an event (A) occurring if another event (B) occurs
  - Event A and event B are NOT mutually exclusive
- Example: (A)heart-attack given (B)chest-pains

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

- Conditional probability is defined in terms of joint events of A and B …

Joint Probability

- *"the number of times A and B can occur"* is called the **joint probability** of A and B or $\mathbf{p(A \cap B)}$
- Therefore, the conditional probability formula is:

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

- Similarly,

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

- where

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

Bayes Rule

From the previous formulas, it follows;

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

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

Substituting into this formula: $$p(A|B) = \frac{p(A \cap B)}{p(B)}$$

Yields the **Bayesian Rule**;

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

Bayesian Rule

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

- P(A|B) is the conditional probability that event A occurring given that event B has occurred.
- P(B|A) is the conditional probability that event B occurring given that even A has occurred.
- P(A) is the probability of event A occurring.
- P(B) is the probability of event B occurring.

Single Evidence, Multiple Hypothesis

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

This is a simple expert system where:

- *E* is the evidence or observation
- *p(H)* is the **prior probability** decided by the expert
- *P(E|H)* is the conditional probability decided by the expert
- *P(H|E)* is the **posterior probability** computed by the expert system using the user supplied input, *E*

Multiple Hypothesis & Evidences

- In reality, there are many hypotheses and evidences- and Bayes Rule will get complicated

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

- N=number of evidence; m=number of hypothesis
- This requires the conditional probabilities of all possible combinations of evidences for all hypothesis! – not a practical option

Naïve Bayes formula

- Therefore, we make the assumption of **conditional independence** among different evidences:

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

- This formula is called **Naïve Bayes** and is more practical and can be implemented in rule-based systems

Application of Naïve Bayes

- We are going to build an expert system with the help of an expert
- The expert is given three conditionally independent evidences E_1, E_2 and E_3 - he creates three mutually exclusive and exhaustive hypotheses H_1, H_2 and H_3, and provides **prior probabilities** for these hypotheses: $p(H_1)$, $p(H_2)$ and $p(H_3)$
- The expert also determines the conditional probabilities of observing each evidence for all possible hypotheses – $P(E|H)$

Conditional & Prior Probabilities

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

Reasoning (1st Iteration)

- Assume that we first observe **evidence *E*$_3$** and inputs into the system. The system computes the posterior probabilities for all hypotheses as

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

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

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

H_2 is marginally more likely

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

Reasoning (2^{nd} Iteration)

- Suppose we now have additional **evidence *E₁***:

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

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

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

H_2 is now even more likely

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

Reasoning (3^{rd} Iteration)

- Finally, we observe **evidence *E₂***:

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

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

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

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

H_3 becomes the most likely

Naïve Bayes Reasoning

- Naïve Bayes reasoning is an **uncertainty handling method** used in Rule-Based Systems
- **Prior probabilities** *p(H)* and **conditional probabilities** *p(E|H)* are required from the expert
- The users will provide **observations** (evidences) and input into the RBS. Rules will then fire and *p(H|E)* probabilities are calculated. When more observations are available (or requested by the RBS), the RBS will fire more rules until a conclusion is reached

Bayesian Spam Filter

$$p(S|W) = \frac{p(W|S) \times p(S)}{p(W|S) \times p(S) + p(W|H) \times p(H)}$$

p(S|W)	probability that a message is a spam, given a word is found in it;
p(S)	overall probability that any given message is spam;
p(W|S)	probability that the word appears in spam messages;
p(H)	overall probability that any given message is not spam (is "ham");
p(W|H)	probability that the word appears in ham messages.

Spam Filter Example

	Ham	Spam	Total
All emails	400	600	1000
"free" in the subject	100	300	400
"viagra" in the subject	10	90	100

	Ham	Spam
P(H)	0.4	0.6
P("free"|H)	0.25	0.5
P("viagra"|H)	0.025	0.15

```
P(spam|"free") = 0.75
P(spam|"viagra") = 0.9
P(spam|"free","viagra") = 0.95
```


Pros and Cons of Bayes Method

- Probability is the oldest and established method for dealing with inexact knowledge and data
- It works well in areas where statistical data is usually available and accurate probability statements can be made –e.g. PROSPECTOR
- However, it is unsuitable in many areas where :
  – reliable statistical information is not available
  – we cannot assume the conditional independence of evidence
  – exhaustive set of hypothesis and evidences are needed
  – experts and users must know the many probabilities
- No way of providing simple explanations during rule firing and how conclusions were reached

Certainty Factors

- A popular alternative to Bayes and widely used
- Developed for MYCIN because experts found it hard to quantify knowledge & facts using probability or mathematical theories
- Certainty factors allow experts to fairly easily express their personal "probability" and, it also allows knowledge engineers to easily incorporate them in rule-based systems
- Certainty Factors are **Measures of Belief** – or how much **confidence** we have in the knowledge/data

Certainty Factors (Cont'd)

- Certainty Factors can be incorporated into Rules and Facts
- CF are typically in this range $-1.0 \leq CF \leq +1.0$
  - ❖ CF = +1.0 the rule/fact is certainly true
  - ❖ CF = 0.0 we don't know whether it is true or not
  - ❖ CF = -1.0 the rule/fact is certainly false

Certainty Factors in Rules

- CF in a rule represents the **expert's confidence or belief** in that **chunk of knowledge**
- Rules with CF has the following structure:

```
IF good_earnings THEN share_up {cf 0.7}
IF win_contract THEN share_up {cf 0.9}
```

- If the condition is true then the conclusion is known to be true (proportionate to the strength of the CF)
- Rules can be elicited by "*How confident are you that good earnings will cause the share price to go up?*"

Certainty Factors in Facts

- CF in facts represents the expert's or user's **belief** in that **piece of information**:

```
good_earnings {cf -0.7}
win_contract {cf 0.8}
```

- Facts can consist of evidence, observations, intuition, etc. and therefore can be subjective
- It can also be based on real probability or obtained through some statistical analysis
- CF can be elicited by "*What are the chances of the company winning the contract?*"

Uncertain Terms Interpretation

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

Referece: Negnevitsky, “Artificial Intelligence”, 2005

Reasoning with Certainty Factors

- Certainty factors are propagated through the reasoning chain when rules are fired
- The following is a typical sequence of CF propagation:

1. User inputs a fact with a certainty value
2. Applicable rules are triggered for firing
3. When a rule is fired, the **net rule certainty** is calculated (slide 33)
4. When many rules are fired, their combined **net conclusion** is calculated (slide 36)
5. Finally a conclusion is given with a certainty

Finding the Net Certainty of a Rule

- When a rule is fired, the **net certainty** of the rule conclusion is calculated as follows:

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

cf(H,E) – Net Certainty of the rule conclusion
cf(E) – Certainty of the fact (rule input)
cf(R) – Certainty of the rule

For example:

IF earnings=good THEN shares=up {cf 0.7}

and the current certainty of earnings=good is 0.8, then

$$cf\ (H,E) = 0.8 \times 0.7 = 0.56$$

This result can be interpreted as "*shares will probably go up*".

Conjunctive Evidences

- For rules with **conjunctive evidences** the certainty of the hypothesis *H* is calculated as follows:

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

For example:

IF earnings=good AND contract=big THEN shares=up {cf 0.9}

current certainty of earnings=good is 0.8, and contract=big is 0.1 then

$$cf\ (H, E_1 \cap E_2) = min\ [0.8, 0.1] \times 0.9 = 0.1 \times 0.9 = 0.09$$

This result can be interpreted as "*it is unknown if shares will go up*"

Disjunctive Evidences

- For rules with **disjunctive evidences** the certainty of the hypothesis *H* is calculated as follows:

$$cf\,(H, E1 \cup E2 \cup \ldots \cup E_n) = max\,[cf\,(E_1), cf\,(E_2), \ldots, cf\,(E_n)] \times cf$$

For example:

IF earnings=good OR contract=big THEN shares=up {cf 0.9}

current certainty of earnings=good is 0.8, and contract=big is 0.1 then

$$cf\,(H, E_1 \cup E_2) = max\,[0.8, 0.1] \times 0.9 = 0.8 \times 0.9 = 0.72$$

This result can be interpreted as "*shares will most probably go up*"

Combining Multiple Conclusions

- When rules are fired, they **assert** their respective ***cf(H,E)*** into working memory
- When the same Hypothesis *H* is asserted by two or more rules, i.e. $cf(H,E_1) \ldots cf(H,E_n)$, all the *cf* are combined to yield a single *cf(H)*

For example:

IF earnings=good THEN shares=up {cf 0.7}
IF contract=big THEN shares=up {cf 0.9}

and earnings=good is 0.8, and contract=big is 0.1 then

$$cf(H,E_1)= \mathbf{0.56} \quad cf(H,E_2)= \mathbf{0.01}$$

- What will be the advice? – "*share will probably go up*" or "*it is unknown*"

Combining Multiple Conclusions (cont)

- When several rules are fired that lead to the same conclusion, we combine them as follows:

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

$\boldsymbol{cf_1} = cf(H,E_1)$ is the net certainty of the conclusion in Rule 1
$\boldsymbol{cf_2} = cf(H,E_2)$ is the net certainty of the conclusion in Rule 2

Certainty Factor – A Complete Example

R1: IF dividends=yes AND
mgnt=good AND
earnings=positive
THEN buy=yes (0.6)

R2: IF contract=large
THEN buy=yes (1.0)

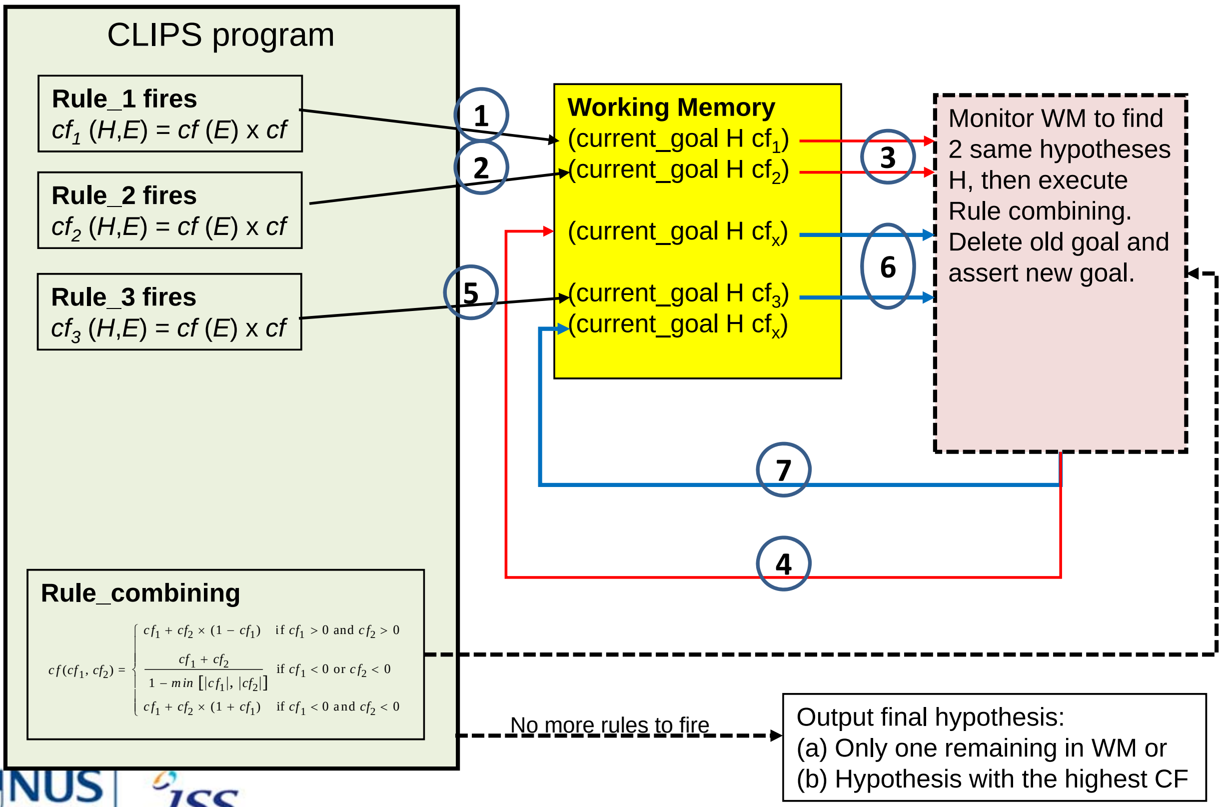
R3: IF stock=penny
THEN buy=yes (-0.7)

Input: dividends=yes (cf 0.9)
mgnt=good (cf 0.7)
earnings=positive (cf 0.5)
contract=large (cf 0.8)
stock=penny (cf 1.0)

```
Fire R1: buy=yes = min (0.9, 0.7, 0.5) * 0.6 = 0.3
Fire R2: buy=yes = 0.8 * 1.0 = 0.8
Fire RX: buy=yes = 0.3 + 0.8 * (1.0 – 0.3) = 0.86
Fire R3: buy=yes = 1.0 * -0.7 = -0.7
Fire RX: buy=yes = (-0.7 + 0.86) / (1.0 – 0.7) = 0.53

Therefore, final recommendation: buy=yes (0.53)
```

Reference Architecture of RBS with CF



Summary

Last

- Certainty factors theory provides a **practical alternative to Bayesian reasoning**
- Certainty Factor theory lacks the mathematical correctness of the probability theory but **does mimic the thinking process of a human expert**
  - where the probabilities are not known or are too difficult or expensive to obtain
- Outperforms Bayesian reasoning in such areas as diagnostics- e.g. MYCIN
- The propagation of CF is of ***Linear*** complexity versus Bayesian's exponential complexity
- CF approach can provide **better explanations** to users