Other formalisms of uncertainty reasoning

Yuqing Tang

Doctoral Program in Computer Science
The Graduate Center
City University of New York
ytang@cs.gc.cuny.edu

December 1, 2010

Outline

Other uncertainty reasoning formalisms

- Default logic
- Certainty factor
- Dempster-Shafer theory
- Fuzzy set

Outline

An introductory example

- A good deal of what we know about a world is 'almost always" true, with a few exceptions. For example,

$$Bird(x) \rightarrow Fly(x)$$

but

$$PENGUIN(x) \rightarrow Bird(x)$$
$$PENGUIN(x) \rightarrow \neg Fly(x)$$
$$\dots$$
$$OSTRICH(x) \rightarrow Bird(x)$$
$$OSTRICH(x) \rightarrow \neg Fly(x)$$

- With classical logic, we will need to revise the always always true rule by

$$Bird(x) \wedge \neg PENGUIN(x) \wedge \dots \wedge \neg OSTRICH(x) \rightarrow Fly(x)$$

Default logic

[Reiter, 1980]

- Default logic combines classical logic with domain-specific inference rules

$$\frac{Bird(x) : Mfly(X)}{fly(X)}$$

where Mp stands for "p is consistent" meaning that $\neg p$ can not be derived.

Definition

A default theory is a pair (W, D) where

- $W \subseteq L$ is a set of input knowledge represented in a language L
- D is a set of default rules of the form

$$\frac{\alpha : M\beta_1, \ldots, M\beta_n}{\gamma}$$

where $a, \beta_1, \ldots, \beta_n, \gamma \in L$, and $n \geq 0$.

Default example

- Given a set of default rules

$$D = \{\frac{Bird(x) : Mfly(X)}{fly(X)}\}$$

  From the facts

$$W = \{Bird(tweety)\}$$

  We can infer that $Fly(tweety)$.

- Later, we obtain new evidences saying that *tweety* is a penguin and penguin can not fly, W is updated into

$$W = \{Bird(tweety), Penguin(tweety), Penguin(X) \rightarrow \neg Fly(tweety)\}$$

  We should be able to derive $\neg Fly(tweety)$.

Inferences in default logic

A set of default rules D inducing an extension E of some underlying knowledge W where

- $E \supseteq W$
- E is deductively closed (no more standard deductive rules can be applied on E to derive new conclusions)
- For all the default rule $\frac{a:M\beta_1,\ldots,M\beta_n}{\gamma} \in D$, if $\alpha \in E$ and $\neg\beta_1 \notin E, \ldots, \neg\beta_n \notin E$, then $\gamma \in E$.

Summary

- Default logic supports reasoning with incomplete information.
- Defaults can be found naturally in many application domains, such as diagnostic problems, information retrieval, legal reasoning, regulations, specifications of systems and software, etc.
- Default Logic enables compact representation of information
- Important prerequisites for successful applications
  - the understanding of the basic concepts, and
  - the existence of powerful implementations.
- The difficulty of understanding Default Logic should not be underestimated.

Outline

Certainty factors

[Heckerman, 1992]

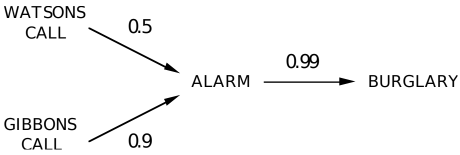
Example

Mr. Holmes receives a telephone call from his neighbor Dr. Watson stating that he hears a burglar alarm sound from the direction of Mr. Holmes house. Preparing to rush home, Mr. Holmes recalls that Dr. Watson is known to be a tasteless practical joker, and he decides to first call his other neighbor, Mrs. Gibbons, who, despite occasional drinking problems, is far more reliable.

A miniature rule-based system for Mr. Holmes situation contains the following rules:

- R_1: if WATSONS CALL then ALARM, $CF_1 = 0.5$
- R_2: if GIBBONS CALL then ALARM, $CF_2 = 0.9$
- R_3: if ALARM then BURGLARY, $CF_3 = 0.99$

The meaning of certainty factors



A rule with certainty factor: IF e THEN h, CF

- A seraintfy factor (CF) represents a person's change in belief in the hypothesis (h) given the evidence (e)
- CF between 0 and 1 means that the person's belief in h given e increases
- CF between -1 and 0 means that the person's belief in h given e decreases

Inferences with certainty factors (Parallel-combination)

To combine R_1 and R_2 into R_4:

- R_1: if WATSONS CALL then ALARM, $CF_1 = 0.5$
- R_2: if GIBBONS CALL then ALARM, $CF_2 = 0.9$
- R_4: if WATSONS CALL and GIBBONS CALL then ALARM, CF_4

$$CF_4 = \begin{cases} CF_1 + CF_2 - CF_1 CF_2, & \text{if } CF_1 \geq 0, CF_2 \geq 0 \\ CF_1 + CF_2 + CF_1 CF_2, & \text{if } CF_1 < 0, CF_2 < 0 \\ \frac{CF_1 + CF_2}{1 - min(|CF_1|, |CF_2|)} & \text{otherwise} \end{cases}$$

$$CF_4 = 0.5 + 0.9 - (0.5)(0.9) = 0.95$$

Inferences with certainty factors (Serial-combination)

To combine R_3 and R_4 into R_5:

- R_3: if ALARM then BURGLARY, $CF_3 = 0.99$
- R_4: if WATSONS CALL and GIBBONS CALL then ALARM, $CF_4 = 0.95$
- R_5: if WATSONS CALL and GIBBONS CALL then BURGLARY, CF_5

$$CF_5 = \begin{cases} CF_3 CF_4, & \text{if } CF_3 > 0 \\ 0 & \text{if } CF_3 \leq 0 \end{cases}$$

$$CF_5 = 0.99 \times 0.95 = 0.94$$

Inferences with certainty factors (Conjunction-combination)

To combine R_6, R_7 and R_8 into R_9:

- R_6: if CHEST PAIN and SHORTNESS OF BREATH then HEART ATTACK, $CF_6 = 0.9$
- R_7: if PATIENT GRIMACES then CHEST PAIN, $CF_7 = 0.7$
- R_8: if PATIENT CLUTCHES THROAT then SHORTNESS OF BREATH, $CF_8 = 0.9$
- R_9: if PATIENT GRIMACES and PATIENT CLUTCHES THROAT then HEART ATTACK, CF_9

$$CF_9 = CF_6 \times min(CF_7, CF_8) = 0.9 \times min(0.7, 0.9) = 0.63$$

Inferences with certainty factors (Disjunction-combination)

To combine R_{10}, R_{11} and R_{12} into R_{13}:

- R_{10}: if h_1 or h_2 then h_3
- R_{11}: if e_1 then h_1
- R_{12}: if e_2 then h_2
- R_{13}: if e_1 or e_2 then h_3

$$CF_{13} = CF_{10} \cdot max(CF_{11}, CF_{12})$$

Problems with certainty factors

- Principle of modularity: Given the logical rule if e then h, and given that e is true, we can assert that h is true
  - (principle of detachment) no matter how we established that e is true, and
  - (principle of locality) no matter what else we know to be true.

  Unfortunately, uncertain reasoning often violates the principles of detachment and locality
- Multiple causes of the same effect: A need to include rules for every possible combinations of observations
- Probabilistic interpretation of certainty factors shows that the combinations functions actually impose assumptions of conditional independence on the propositions involved in the combinations
  - This is usually violated in applications
- CF model requires that we encode the rules in the direction in they are applied
  - This is usually counter-intuitive for the experts
- The authors of [Heckerman, 1992] suggested that Bayesian network and its variants are remedies for the above problems

Outline

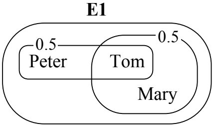
An introductory example

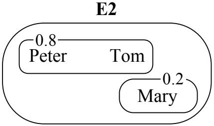
[Haenni, 2002]

- Evidence 1:
  Mr. Jones was assassinated by the Mafia. An informer tells the police that the selection of the assassin was done as follows:
  - a fair coin is tossed
  - "head" $\rightarrow$ either Peter or Tom is selected
  - "tail" either Tom or Mary is selected
- Evidence 2:
  The police finds the assassin's fingerprint. An expert states that it is male with 80% chance and female with 20% chance.

Dempster-Shafer Theory vs. Probability theory

Using Dempster-Shafer theory:

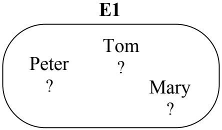


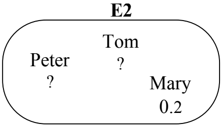


$m_1(\{Peter, Tom\}) = 0.5$
$m_1(\{Tom, Mary\}) = 0.5$

$m_2(\{Peter, Tom\}) = 0.8$
$m_2(\{Mary\}) = 0.2$

Using probability theory:





$P_1(Peter) = P_1(Mary) = 0.25$
$P_1(Tom) = 0.5$

$P_2(Peter) = P_2(Tom) = 0.4$
$P_2(Mary) = 0.2$

Dempster-Shafer Theory

[Shafer, 1976]

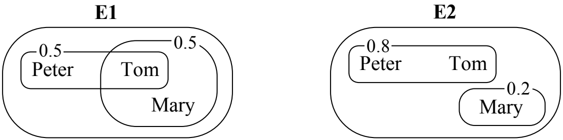
- Frame of Discernment: Θ (set of possible events, one of them is true)
- Multi-Variable Frames:

$$D = \{x_1, \ldots, x_n\} \Rightarrow \Theta_D = \Theta_{x_1} \times \ldots \times \Theta_{x_n}$$

- Belief mass function: $m : 2^{\Theta_D} \to [0,1]$ with $\Sigma_{A \subseteq \Theta_D} m(A) = 1$
- Belief: $Bel(A) = \Sigma_{B \subseteq A} m(A)$
- Focal Sets: $A \subseteq \Theta_D$ s.t. $m(A) \neq 0$
- Dempsters rule of combination:

$$m_{1,2}(A) = \frac{\Sigma_{B \cap C = A} m_1(B) \cdot m_2(C)}{\Sigma_{B \cap C \neq \emptyset} m_1(B) \cdot m_2(C)}$$

Combination example

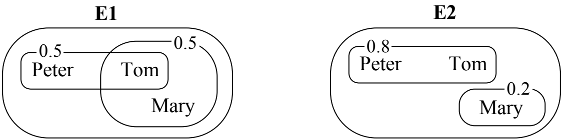


$$m_1(\{Peter, Tom\}) = 0.5 \qquad m_2(\{Peter, Tom\}) = 0.8$$
$$m_1(\{Tom, Mary\}) = 0.5 \qquad m_2(\{Mary\}) = 0.2$$

$$\begin{aligned} n &= m_1(\{P,T\}) \cdot m_2(\{P,T\} + m_1(\{T,M\}) \cdot m_2(\{P,T\}) \\ &\quad + m_1(\{T,M\}) \cdot m_2(\{M\}) \\ &= 0.5 \cdot 0.8 + 0.5 \cdot 0.8 + 0.5 \cdot 0.2 = 0.9 \end{aligned}$$

$$\begin{aligned} & m(\{Peter, Tom\}) \\ &= \frac{m_1(\{P,T\}) \cdot m_2(\{P,T\})}{n} = \frac{0.5 \cdot 0.8}{0.9} = 0.44 \end{aligned}$$

Combination example (cont.)



$$m_1(\{Peter, Tom\}) = 0.5 \qquad m_2(\{Peter, Tom\}) = 0.8$$
$$m_1(\{Tom, Mary\}) = 0.5 \qquad m_2(\{Mary\}) = 0.2$$

$$\begin{aligned} & m(\{Tom\}) \\ = \; & \frac{m_1(\{T, M\}) \cdot m_2(\{P, T\})}{n} = \frac{0.5 \cdot 0.8}{0.9} = 0.44 \\ & m(\{Mary\}) \\ = \; & \frac{m_1(\{T, M\}) \cdot m_2(\{M\})}{n} = \frac{0.5 \cdot 0.2}{0.9} = 0.11 \end{aligned}$$

Combination example (cont. II)

$$
\begin{aligned}
m(\{Peter, Tom\}) &= 0.44 \\
m(\{Tom\}) &= 0.44 \\
m(\{Mary\}) &= 0.11
\end{aligned}
$$

$$
\begin{aligned}
Bel(Peter) &= 0.44 \\
Bel(Tom) &= 0.44 + 0.44 = 0.88 \\
Bel(Mary) &= 0.11
\end{aligned}
$$

Summary I

Benefits of Dempster-Shafer Theory:

- Allows proper distinction between reasoning and decision taking
- No modeling restrictions (e.g. DAGs)
- It represents properly partial and total ignorance
- Ignorance is quantified:
  - low degree of ignorance means
    - high confidence in results
    - enough information available for taking decisions
  - high degree of ignorance means
    - low confidence in results
    - gather more information (if possible) before taking decisions
- Conflict is quantified:
  - low conflict indicates the presence of confirming information sources
  - high conflict indicates the presence of contradicting sources
- Simplicity: Dempsters rule of combination covers
  - combination of evidence,

Summary II

- Bayes rule,
- Bayesian updating (conditioning),
- belief revision (results from non-monotonicity),
- etc.

DS-Theory is not very successful because:

- Inference is less efficient than Bayesian inference

[Haenni, 2002]

Outline

An introductory example

Example

"Hans ate X eggs for breakfast", where $X \in U = \{1, 2, ..., 8\}$.

$$
\begin{array}{lcl}
U & = & [\ 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8\ ] \\
p & = & [\ .1 \quad .8 \quad .1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0\ ] \\
\pi & = & [\ 1 \quad 1 \quad 1 \quad 1 \quad .8 \quad .6 \quad .4 \quad .2\ ]
\end{array}
$$

where p is the probability distribution and π is a fuzzy possibility distribution.

The possiblity for $X = 3$ is 1, the probability is only 0.1.

A logic based on the two truth values *True* and *False* is sometimes inadequate when describing human reasoning. The logic of fuzzy set uses the whole interval between 0 (*False*) and 1 (*True*) to describe human reasoning.

Fuzzy sets

[Zadeh, 1965]

- A set A is in terms of its characteristic function

$$\mu_A(x) : U \to [0,1]$$

  A point x belongs to set A with possibility $_A(x)$.

- Union

$$\mu_{A \cup B}(x) = max(\mu_A(x), \mu_B(x))$$

- Intersection

$$\mu_{A \cap B}(x) = min(\mu_A(x), \mu_B(x))$$

- Complement

$$\mu_{\bar{A}}(x) = 1 - \mu_A(x)$$

Buying house example

Example

A four-person family want sto buy a house. The indication of "comfortable" (**c**) is the number of bedrooms, and the concept of large (**l**).

$$
\begin{array}{lcllllllllllll}
\mathbf{U} & = & [& 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10\ ] \\
\mathbf{c} & = & [& .2 & .5 & .8 & 1 & .7 & .3 & 0 & 0 & 0 & 0\ ] \\
\mathbf{l} & = & [& 0 & 0 & .2 & .4 & .6 & .8 & 1 & 1 & 1 & 1\ ] \\
\mathbf{c} \cap \mathbf{l} & = & [& 0 & 0 & .2 & .4 & .6 & .3 & 0 & 0 & 0 & 0\ ] \\
\mathbf{c} \cup \mathbf{l} & = & [& .2 & .5 & .8 & 1 & .7 & .8 & 1 & 1 & 1 & 1\ ] \\
\bar{\mathbf{l}} & = & [& 1 & 1 & .8 & .6 & .4 & .2 & 0 & 0 & 0 & 0\ ]
\end{array}
$$

Fuzzy relations

$$(a, b) \in R \subseteq A \times B$$

- $\mu_R(a, b) : A \times B \to [0, 1]$
- Composition $\mu_{R_1 \times R_2}(a, c) = max_b\left(min\left(\mu_{R_1}(a, b), \mu_{R_2}(b, c)\right)\right)$

Example

R_1: Huey resembles Dewey and Louie

	Dewey	Louie
Huey	0.8	0.9

R_2: Dewey and Louie resembles Donald

	Donald
Dewey	0.5
Louie	0.6

- Huey resembles (0.8) Dewey, and Dewey resembles (0.5) Donald, or
- Huey resembles (0.9) Louie, and Louie resembles (0.6) Donald

Huey resembles Donald:

$$max(min(0.8, 0.5), min(0.9, 0.6) = max(0.5, 0.6) = 0.6$$

Summary

- Applications
  - Expert systems: decision-support systems, financial planners, diagnostic systems
  - Information retrieval systems
  - Control: a navigation system for automatic cars, a predicative fuzzy-logic controller for automatic operation of trains, laboratory water level controllers, controllers for robot arc-welders, feature-definition controllers for robot vision, graphics controllers for automated police sketchers, and more
- Fuzzy systems, including fuzzy logic and fuzzy set theory, provide a rich and meaningful addition to standard logic.
- The mathematics behind fuzzy theories is consistent.
- Fuzzy theories provide the opportunity for modeling of conditions which are inherently imprecisely defined.

Outline

Summary

- Default logic
- Certainty factor
- Dempster-Shafer theory
- Fuzzy set

References I

Rolf Haenni.
Introduction to dempster-shafer theory.
Weekly Seminar, PPM Group, University of Konstanz, Germany, November 2002.

David Heckerman.
The certainty-factor model.
In Steven Shapiro, editor, *Encyclopedia of Artificial Intelligence*, pages 131–138. Wiley, New York, second edition edition, 1992.

Raymond Reiter.
A logic for default reasoning.
Artificial Intelligence, 13(1-2):81–132, 1980.

G. Shafer.
A mathematical theory of evidence.
Princeton university press, 1976.

References II

L.A. Zadeh.
Fuzzy sets.
Information Control, 8:338–353, 1965.