Chapter Overview

Inexact Reasoning

Introduction

Sources of Uncertainty

Bayesian Approaches

Certainty Factors

Dempster-Shafer Theory of Evidence

Fuzzy Logic

State of Inexact Reasoning

Chapter Review

Introduction

reasoning under uncertainty and with inexact knowledge

heuristics

ways to mimic heuristic knowledge processing methods of experts

empirical associations

based on limited observations

probabilities

objective (frequency counting) vs. subjective (human-oriented)

reproduceability

in case of doubt, can the observations be repeated, and will they deliver the same result?

approaches

numerical

symbol-oriented

constraints

computer time

memory space

money (e.g. for data collection)

exact reasoning, in contrast to inexact reasoning, is based on *deductive* approaches to logic

Objectives

for dealing with uncertainty and inexact knowledge

expressiveness

Can concepts used by humans be represented adequately? Can the confidence of experts in their choices be expressed?

comprehensibility

How difficult is it to understand the representation and the evaluation?

soundness

Are probability laws required (sum of conditional probabilities = 1), or is a relevance ranking sufficient?

consistency

similar results for similar inputs

example data

Are large quantities of historic data needed?

reasoning

long inference chains

computational complexity

Are the required calculations feasible?

portability

Can the method be used with any system and application?

Sources of Uncertainty

and inexact information

data

missing data

unreliable, ambiguous

inconsistent or imprecise representation of data

skewed by the user's perception ("best guess")

derived from defaults

expert knowledge

inconsistency between different experts

plausible: "best guess" of the expert

statistical associations observed by the expert

limited applicability

knowledge representation

represented knowledge doesn't exactly model the real system

inference process

- deductive: the application of a rule is formally correct, but the result is wrong
- inductive: new conclusions are obtained in an uncertain or inexact way

Uncertainty and Rules

individual rules

- errors (see previous transparency)
- likelihood of evidence
  for each premise, and for the conclusion
- combination of evidence
  from multiple premises

conflict resolution

- explicit priority provided by the expert
  relative or absolute ranking of rules
- implicit priority derived from rule properties
  specificity of patterns
  recency of facts matching patterns
  ordering of patterns (lexicographic, means end)
  order that rules are entered

compatibility

- contradiction of rules
- subsumption
- redundancy
- missing rules
- data fusion

Basics of Probability Theory

mathematical approach for processing uncertain information

sample space set $X = \{x_1, x_2, \ldots, x_n\}$

collection of all possible events

can be discrete or continuous

probability number $P(x_i)$

likelihood of an event x_i to occur

- non-negative value from [0,1]
- the total probability of the entire sample space is 1
- for mutually exclusive events, the probability that at least one of them will occur is the sum of their individual probabilities
- *experimental* probability (a posteriori) based on the frequency of events
- *subjective* probability based on estimates of experts

types of probabilities

- *a priori* probability (classical, theoretical)
  repeatable events
  can be calculated exactly

  $$P(E) = \frac{W}{N}$$

  where W is the number of outcomes of E for N possible outcomes

- *a posteriori* probability (experimental)
  repeatable events
  approximated from experiments

  $$P(E) = \lim_{N \to \infty} \frac{f(E)}{N}$$

  where $f(E)$ is the frequency that E is observed for N possible outcomes

- *subjective* probability (personal)
  non-repeatable events
  no calculation or approximation available
  based on expert's judgement

compound probabilities

- for *independent* events that do not affect each other in any way
- *joint* probability (intersection) of two independent events A and B

$$P(A \cap B) = \frac{n(A \cap B)}{n(S)} = P(A) \times P(B)$$

  where $n(S)$ is the number of elements in S

- *union* probability of two independent events A and B

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= P(A) + P(B) - P(A) \times P(B) \end{aligned}$$

conditional probabilities

- for *dependent* events that affect each other in some way
- *conditional* probability of event A given that event B has already occurred

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

Advantages and Problems

+ formal foundation

+ reflection of reality (a posteriori)

- may be inappropriate (future different from past)

- inexact or incorrect (subjective probabilities)

- some knowledge is represented implicitly

Bayesian Approaches

derive the probability of a cause given a symptom

inverse probability also a posteriori probability

inverse to *conditional* probability

probability of an earlier event given that a later one occurred

Bayes' rule single hypothesis, single event

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

or

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E|H) * P(H) + P(E|\neg H) * P(\neg H)}$$

where H is a hypothesis, and E an event

Bayes' rule multiple hypotheses, multiple events

posterior probability of hypothesis H_i from evidence $E_1, \ldots, E_n$

$$P(H_i|E_1, E_2, \ldots E_n) = \frac{P(E_1, E_2, \ldots E_n|H_i) * P(H_i)}{p(E_1, E_2, \ldots E_n)}$$

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

where the pieces of evidence E_i are independent

Advantages and Problems

+ sound theoretical foundation

+ well-defined semantics for decision making

- require large amounts of probability data

- independence assumption of evidences frequently not valid

- sources of prior and conditional probabilities: statistics (sufficient sample sizes), human experts (consistent, comprehensive, trustworthy)

- relationship between hypothesis and evidence is reduced to a number

- explanations difficult to provide to the user

Example

[Gonzalez and Dankel, 1993], p. 235-238

`IF the patient has a cold THEN the patient will sneeze (0.75)`

Given:

$$\begin{aligned} P(H) &= P(\text{Rob has a cold}) &= 0.2 \\ P(E|H) &= P(\text{Rob sneezed}|\text{Rob has a cold}) &= 0.75 \\ P(E|\neg H) &= P(\text{Rob sneezed}|\text{Rob has no cold}) &= 0.2 \end{aligned}$$

Then

$$\begin{aligned} P(E) &= P(\text{Rob sneezed}) \\ &= (0.75 * 0.2) + (0.2 * 0.8) \\ &= 0.15 + 0.16 = 0.31 \\ P(H|E) &= P(\text{Rob has a cold}|\text{Rob sneezed}) \\ &= \frac{0.75 * 0.2}{0.31} \\ &= 0.48387 \\ P(H|\neg E) &= P(\text{Rob has a cold}|\text{Rob didn't sneeze}) \\ &= \frac{P(\neg E|H) * P(H)}{p(\neg E)} \\ &= \frac{(1 - 0.75) * 0.2}{1 - 0.31} \\ &= 0.07246 \end{aligned}$$

Certainty Factors

alternative to Bayesian methods

basic idea

denotes the belief in a hypothesis h given that some pieces of evidence E_i are observed

does not make any statement about the belief if no evidence is present (in contrast to Bayesian methods)

certainty factor

$$CF = \frac{MB - MD}{1 - min(MB, MD)}$$

CF ranges between -1 (denial of h) and 1 (confirmation of h)

measure of belief

degree to which hypothesis h is supported by evidence e

$$MB(H,E) = \begin{cases} 1 & if P(H) = 1 \\ \frac{P(H|E) - P(H)}{1 - P(H)} & otherwise \end{cases}$$

measure of disbelief

degree to which doubt in hypothesis h is supported by evidence e

$$MD(H,E) = \begin{cases} 1 & if P(H) = 0 \\ \frac{P(H) - P(H|E)}{P(H)} & otherwise \end{cases}$$

combining antecedent evidence

use of premises with less than absolute confidence

$E_1 \wedge E_2$	$min(CF(H,E_1), CF(H,E_1))$
$E_1 \vee E_2$	$min(CF(H,E_1), CF(H,E_1))$
$\neg E$	$-CF(H,E)$

combining certainty factors for the same conclusion

several rules can be used to come to the same conclusion

applied incrementally as new evidence becomes available

$$CF_{rev}(CF_{old}, CF_{new})$$
$$= \begin{cases} CF_{old} + CF_{new}(1 - CF_{old}) & both > 0 \\ CF_{old} + CF_{new}(1 + CF_{old}) & both < 0 \\ \frac{CF_{old}+CF_{new}}{1-min(|CF_{old}|,|CF_{new}|)} & one < 0 \end{cases}$$

Advantages and Problems

+ simple implementation

+ better modeling of human experts' beliefs
  expression of belief and disbelief

+ successful application for certain problem classes

+ easier to gather than other values (no statistical
  base required

- (partially) ad hoc approach

- combination of non-independent evidence
  unsatisfactory

- new knowledge may require changes in the CFs of
  existing knowledge

- certainty factors can become the opposite of
  conditional probabilities for certain cases

- not suitable for long inference chains

Dempster-Shafer Theory

mathematical theory of evidence

frame of discernment FD

power set of the set of possible conclusions

mass probability function m

assigns a value from [0,1] to every item in the frame of discernment

mass probability $m(A)$

portion of the total mass probability that is assigned to an element A of FD

it cannot be further subdivided

belief $Bel(A)$ in a subset A

sum of the mass probabilities of all the proper subsets of A

likelihood that one of its members is the conclusion

certainty $Cer(A)$

interval $[Bel(A) \quad Pl(A)]$ indicating the range of belief

$Pl(A)$ is the plausibility, or maximum belief

combination of mass probabilities

$$m_1 \oplus m_2(C) = \frac{\sum_{X \cap Y = C} m_1(X) * m_2(Y)}{1 - \sum_{X \cap Y = \emptyset} m_1(X) * m_2(Y)}$$

where X, Y are hypothesis subsets Y and C is their intersection

Advantages and Problems

+ clear, rigorous foundation

+ able to express confidence ("certainty about certainty") through intervals

- non-intuitive determination of mass probability values

- usability somewhat unclear

- high computational overhead

- normalization may lead to counterintuitive results

Fuzzy Logic

linguistic variable

natural language term to describe concepts with vague values

fuzzy set

the categorization of elements x_i into a set S is described through a membership function $\mu_S(x)$ that associates each element x_i with a degree of membership in S

possibility measure $Poss\{x \in S\}$

degree to which an individual element x is a potential member in a fuzzy set S

possibility refers to allowed values

probability expresses expected occurrences of events

multiple premises

$$Poss(A \wedge B) = min(Poss(A), Poss(B))$$
$$Poss(A \vee B) = max(Poss(A), Poss(B))$$

fuzzy inference

$$Poss(B|A) = min(1, (1 - Poss(A) + Poss(B)))$$

implication according to MAX-MIN inference
also MAX-PRODUCT inference and others

Advantages and Problems

+ rather general theory of uncertainty
+ wide applicability, many applications
+ natural use of vague and imprecise concepts

- membership functions can be difficult to find
- problems with long inference chains

State of Inexact Reasoning

no single best method

not even agreement on the measurement criteria

relations between approaches

is Dempster-Shafer a generalization of classical probability, or the other way round?

Bayesian networks

used for the integration of uncertainty in knowledge-based systems

computational complexity

extremely high (exponential) for initial Dempster-Shafer approaches, refined approaches are more feasible

fuzzy logic

widely used for control applications

revitalization of probability theory

re-examination of its foundations and usability

Chapter Review

Inexact Reasoning

Introduction

probability, heuristics; expressiveness, soundness, consistency

Sources of Uncertainty

data, expert knowledge, inference process

Bayesian Approaches

derive the probability of a cause given a symptom; uses Bayes' rule

Certainty Factors

belief in a hypothesis given that some pieces of evidence are observe

Dempster-Shafer Theory of Evidence

mathematical theory of evidence based on intervals

Fuzzy Logic

uses natural language terms to describe concepts with vague values

State of Inexact Reasoning

wide use of fuzzy logic in niche applications,
some use of other approaches

Chapter Review