# CPE/CSC 481: Knowledge-Based Systems

*Franz J. Kurfess*

*Computer Science Department*
*California Polytechnic State University*
*San Luis Obispo, CA, U.S.A.*

# Usage of the Slides

- these slides are intended for the students of my CPE/CSC 481 “Knowledge-Based Systems” class at Cal Poly SLO
  - if you want to use them outside of my class, please let me know (fkurfess@calpoly.edu)
- I usually put together a subset for each quarter as a “Custom Show”
  - to view these, go to “Slide Show => Custom Shows”, select the respective quarter, and click on “Show”
  - in Apple Keynote, I use the “Hide” feature to achieve similar results
- To print them, I suggest to use the “Handout” option
  - 4, 6, or 9 per page works fine
  - Black & White should be fine; there are few diagrams where color is important

# Overview Reasoning and Uncertainty

- Motivation
- Objectives
- Sources of Uncertainty and Inexactness in Reasoning
  - Incorrect and Incomplete Knowledge
  - Ambiguities
  - Belief and Ignorance
- Probability Theory
  - Bayesian Networks
  - Certainty Factors
  - Belief and Disbelief
  - Dempster-Shafer Theory
  - Evidential Reasoning
- Important Concepts and Terms
- Chapter Summary

# Motivation

- reasoning for real-world problems involves missing knowledge, inexact knowledge, inconsistent facts or rules, and other sources of uncertainty
- while traditional logic in principle is capable of capturing and expressing these aspects, it is not very intuitive or practical
  - explicit introduction of predicates or functions
- many expert systems have mechanisms to deal with uncertainty
  - sometimes introduced as ad-hoc measures, lacking a sound foundation

# Objectives

- be familiar with various sources of uncertainty and imprecision in knowledge representation and reasoning
- understand the main approaches to dealing with uncertainty
  - probability theory
    - Bayesian networks
    - Dempster-Shafer theory
  - important characteristics of the approaches
    - differences between methods, advantages, disadvantages, performance, typical scenarios
- evaluate the suitability of those approaches
  - application of methods to scenarios or tasks
- apply selected approaches to simple problems

# Introduction

- reasoning under uncertainty and with inexact knowledge
  - frequently necessary for real-world problems
- heuristics
  - ways to mimic heuristic knowledge processing
  - methods used by experts
- empirical associations
  - experiential reasoning
  - based on limited observations
- probabilities
  - objective (frequency counting)
  - subjective (human experience )
- reproducibility
  - will observations deliver the same results when repeated

# Dealing with Uncertainty

- expressiveness
  - can concepts used by humans be represented adequately?
  - can the confidence of experts in their decisions be expressed?
- comprehensibility
  - representation of uncertainty
  - utilization in reasoning methods
- correctness
  - probabilities
    - adherence to the formal aspects of probability theory
  - relevance ranking
    - probabilities don't add up to 1, but the "most likely" result is sufficient
  - long inference chains
    - tend to result in extreme (0,1) or not very useful (0.5) results
- computational complexity
  - feasibility of calculations for practical purposes

# Sources of Uncertainty

- data
  - data missing, unreliable, ambiguous,
  - representation imprecise, inconsistent, subjective, derived from defaults, …
- expert knowledge
  - inconsistency between different experts
  - plausibility
    - “best guess” of experts
  - quality
    - causal knowledge
      - deep understanding
    - statistical associations
      - observations
  - scope
    - only current domain, or more general

# Sources of Uncertainty (cont.)

- knowledge representation
  - restricted model of the real system
  - limited expressiveness of the representation mechanism
- inference process
  - deductive
    - the derived result is formally correct, but inappropriate
    - derivation of the result may take very long
  - inductive
    - new conclusions are not well-founded
      - not enough samples
      - samples are not representative
  - unsound reasoning methods
    - induction, non-monotonic, default reasoning, "common sense"

# Uncertainty in Individual Rules

- errors
  - domain errors
  - representation errors
  - inappropriate application of the rule
- likelihood of evidence
  - for each premise
  - for the conclusion
  - combination of evidence from multiple premises

# Uncertainty and Multiple Rules

- conflict resolution
  - if multiple rules are applicable, which one is selected
    - explicit priorities, provided by domain experts
    - implicit priorities derived from rule properties
      - specificity of patterns, ordering of patterns  creation time of rules, most recent usage, …
- compatibility
  - contradictions between rules
  - subsumption
    - one rule is a more general version of another one
  - redundancy
  - missing rules
  - data fusion
    - integration of data from multiple sources

# 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 $P(x_i)$ reflects the likelihood of an event $x_i$ to occur
  - non-negative value in [0,1]
  - total probability of the sample space (sum of probabilities) is 1
  - for mutually exclusive events, the probability for at least one of them is the sum of their individual probabilities
  - experimental probability
    - based on the frequency of events
  - subjective probability
    - based on expert assessment

# Compound Probabilities

- describes independent events
  - do not affect each other in any way
- joint probability of two independent events A, B

$$P(A \cap B) = n(A \cap B) / n(s) = P(A) * P(B)$$

  - where n(S) is the number of elements in S
- union probability of two independent events A, B

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$= P(A) + P(B) - P(A) * P(B)$$

# Conditional Probabilities

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

$$P(A|B) = P(A \cap B) / P(B)$$

# Advantages and Problems: Probabilities

- advantages
  - formal foundation
  - reflection of reality (a posteriori)
- problems
  - may be inappropriate
    - the future is not always similar to the past
  - inexact or incorrect
    - especially for subjective probabilities
  - ignorance
    - probabilities must be assigned even if no information is available
      - assigns an equal amount of probability to all such items
  - non-local reasoning
    - requires the consideration of all available evidence, not only from the rules currently under consideration
  - no compositionality
    - complex statements with conditional dependencies can not be decomposed into independent parts

# Bayesian Approaches

- derive the probability of a cause given a symptom
- has gained importance recently due to advances in efficiency
  - more computational power available
  - better methods
- especially useful in diagnostic systems
  - medicine, computer help systems
- inverse probability
  - inverse to conditional probability of an earlier event given that a later one occurred

# Bayes' Rule for Single Event

- single hypothesis H, single event E

  $P(H|E) = (P(E|H) * P(H)) / P(E)$

  or

- $P(H|E) = (P(E|H) * P(H) / (P(E|H) * P(H) + P(E|\neg H) * P(\neg H) )$

# Bayes' Rule for Multiple Events

- multiple hypotheses Hi, multiple events E1, …, En

  P(Hi|E1, E2, …, En)
  = (P(E1, E2, …, En|Hi) * P(Hi)) / P(E1, E2, …, En)

  or

  P(Hi|E1, E2, …, En)
  = (P(E1|Hi) * P(E2|Hi) * …* P(En|Hi) * P(Hi)) /
  Σk P(E1|Hk) * P(E2|Hk) * … * P(En|Hk)* P(Hk)

  - with independent pieces of evidence Ei

# Using Bayesian Reasoning: Spam Filters

- Bayesian reasoning was used for early approaches to spam filtering

# Advantages and Problems of Bayesian Reasoning

- advantages
  - sound theoretical foundation
  - well-defined semantics for decision making
- problems
  - requires large amounts of probability data
    - sufficient sample sizes
  - subjective evidence may not be reliable
  - independence of evidences assumption often not valid
  - relationship between hypothesis and evidence is reduced to a number
  - explanations for the user difficult
  - high computational overhead

# Certainty Factors

- denotes the belief in a hypothesis H given that some pieces of evidence E are observed
- no statements about the belief means that no evidence is present
  - in contrast to probabilities, Bayes' method
- works reasonably well with partial evidence
  - separation of belief, disbelief, ignorance
- shares some foundations with Dempster-Shafer (DS) theory, but is more practical
  - introduced in an ad-hoc way in MYCIN
  - later mapped to DS theory

# Belief and Disbelief

- measure of belief
  - degree to which hypothesis H is supported by evidence E
  - MB(H,E) = 1 if P(H) = 1
    (P(H|E) - P(H)) / (1- P(H)) otherwise
- measure of disbelief
  - degree to which doubt in hypothesis H is supported by evidence E
  - MD(H,E) = 1 if P(H) = 0
    (P(H) - P(H|E)) / P(H)) otherwise

# Certainty Factor

- certainty factor CF
  - ranges between -1 (denial of the hypothesis H) and +1 (confirmation of H)
  - allows the ranking of hypotheses
- difference between belief and disbelief

  $$CF(H,E) = MB(H,E) - MD(H,E)$$

- combining antecedent evidence
  - use of premises with less than absolute confidence
    - $E_1 \wedge E_2 = \min(CF(H, E_1), CF(H, E_2))$
    - $E_1 \vee E_2 = \max(CF(H, E_1), CF(H, E_2))$
    - $\neg E = \neg\, CF(H, E)$

# Combining Certainty Factors

- certainty factors that support the same conclusion
- several rules can lead to the same conclusion
- applied incrementally as new evidence becomes available

$CF_{rev}(CF_{old}, CF_{new}) =$

$CF_{old} + CF_{new}(1 - CF_{old})$ if both > 0

$CF_{old} + CF_{new}(1 + CF_{old})$ if both < 0

$CF_{old} + CF_{new} / (1 - \min(|CF_{old}|, |CF_{new}|))$ if one < 0

# Characteristics of Certainty Factors

| Aspect | Probability | MB | MD | CF |
|---|---|---|---|---|
| Certainly true | $P(H\|E) = 1$ | 1 | 0 | 1 |
| Certainly false | $P(\neg H\|E) = 1$ | 0 | 1 | -1 |
| No evidence | $P(H\|E) = P(H)$ | 0 | 0 | 0 |

- Ranges
  - measure of belief $0 \leq MB \leq 1$
  - measure of disbelief $0 \leq MD \leq 1$
  - certainty factor $-1 \leq CF \leq +1$

# Advantages and Problems of Certainty Factors

- Advantages
  - simple implementation
  - reasonable modeling of human experts' belief
    - expression of belief and disbelief
  - successful applications for certain problem classes
  - evidence relatively easy to gather
    - no statistical base required
- Problems
  - partially ad hoc approach
    - theoretical foundation through Dempster-Shafer theory was developed later
  - combination of non-independent evidence unsatisfactory
  - new knowledge may require changes in the certainty factors 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
  - uncertainty is modeled through a range of probabilities
    - instead of a single number indicating a probability
  - sound theoretical foundation
  - allows distinction between belief, disbelief, ignorance (non-belief)
  - certainty factors are a special case of DS theory

# DS Theory Notation

- environment $\Theta = \{O_1, O_2, ..., O_n\}$
  - set of objects $O_i$ that are of interest
  - $\Theta = \{O_1, O_2, ..., O_n\}$
- frame of discernment FD
  - an environment whose elements may be possible answers
  - only one answer is the correct one
- mass probability function m
  - assigns a value from [0,1] to every item in the frame of discernment
  - describes the degree of belief in analogy to the mass of a physical object
- mass probability m(A)
  - portion of the total mass probability that is assigned to a specific element A of FD

# Belief and Certainty

- belief Bel(A) in a set A
  - sum of the mass probabilities of all the proper subsets of A
    - all the mass that supports A
  - likelihood that one of its members is the conclusion
  - also called *support function*
- plausibility Pls(A)
  - maximum belief of A
  - upper bound for the range of belief
- certainty Cer(A)
  - interval [Bel(A), Pls(A)]
    - also called *evidential interval*
  - expresses the range of belief

# Combination of Mass Probabilities

- combining two masses in such a way that the new mass represents a consensus of the contributing pieces of evidence
  - set intersection puts the emphasis on common elements of evidence, rather than conflicting evidence

$$m_1 \oplus m_2 (C) = \Sigma\, X \cap Y\ m_1(X) * m_2(Y)$$

$$= C\ m_1(X) * m_2(Y) / (1- \Sigma X \cap Y)$$

$$= C\ m_1(X) * m_2(Y)$$

where X, Y are hypothesis subsets

C is their intersection $C = X \cap Y$

$\oplus$ is the orthogonal or direct sum

# Differences Probabilities - DS Theory

| *Aspect* | *Probabilities* | *Dempster-Shafer* |
|---|---|---|
| Aggregate Sum | $\sum_i P_i = 1$ | $m(\Theta) \leq 1$ |
| Subset $X \subseteq Y$ | $P(X) \leq P(Y)$ | $m(X) > m(Y)$ allowed |
| relationship X, ¬X (ignorance) | $P(X) + P(\neg X) = 1$ | $m(X) + m(\neg X) \leq 1$ |

# Evidential Reasoning

- extension of DS theory that deals with uncertain, imprecise, and possibly inaccurate knowledge
- also uses evidential intervals to express the confidence in a statement
  - lower bound is called support (Spt) in evidential reasoning, and belief (Bel) in Dempster-Shafer theory
  - upper bound is plausibility (Pls)

# Evidential Intervals

| Meaning | Evidential Interval |
|---|---|
| Completely true | [1,1] |
| Completely false | [0,0] |
| Completely ignorant | [0,1] |
| Tends to support | [Bel,1] where $0 < Bel < 1$ |
| Tends to refute | [0,Pls] where $0 < Pls < 1$ |
| Tends to both support and refute | [Bel,Pls] where $0 < Bel \leq Pls < 1$ |

Bel: belief; lower bound of the evidential interval

Pls: plausibility; upper bound

# Advantages and Problems of Dempster-Shafer

- advantages
  - clear, rigorous foundation
  - ability to express confidence through intervals
    - certainty about certainty
  - proper treatment of ignorance
- problems
  - non-intuitive determination of mass probability
  - very high computational overhead
  - may produce counterintuitive results due to normalization
  - usability somewhat unclear

# Post-Test

# Important Concepts and Terms

- Bayesian networks
- belief
- certainty factor
- compound probability
- conditional probability
- Dempster-Shafer theory
- disbelief
- evidential reasoning
- inference
- inference mechanism
- ignorance
- knowledge
- knowledge representation
- mass function
- probability
- reasoning
- rule
- sample
- set
- uncertainty

# Summary Reasoning and Uncertainty

- many practical tasks require reasoning under uncertainty
  - missing, inexact, inconsistent knowledge
- variations of probability theory are often combined with rule-based approaches
  - works reasonably well for many practical problems
- Bayesian networks have gained some prominence
  - improved methods, sufficient computational power