INF5390 - Kunstig intelligens

Probabilistic Reasoning

Roar Fjellheim

Outline

- Agents and uncertainty
- Probability theory
- Bayes’ rule
- Bayesian networks
- Bayesian inference
- Other approaches
- Summary

AIMA Chapter 13: Quantifying Uncertainty
AIMA Chapter 14: Probabilistic Reasoning

Uncertain knowledge

- In all real domains, the agent must be able to handle *uncertainty*, due to
  - √ *Limited resources*: Cannot exhaustively enumerate all possible situations and consequences of actions
  - √ *Theoretical ignorance*: No theory for the domain exists
  - √ *Practical ignorance*: All necessary data is not available
- The agent still has to *act*, and needs to make decisions where uncertainty is explicitly recognized
- *Probability* can be used to summarize the state of the agent's beliefs (or ignorance)

Status of probability sentences

- Statement
  - “The patient has cavity with probability 0.8”
- In *logic*, a sentence is true or false, depending on the interpretation and the world
- In *probability theory*, the probability assigned to a sentence depends on the *evidence* so far
  - *Prior* (unconditional) probability: Before any evidence
  - *Posterior* (conditional) probability: After some evidence
- Probability is more like entailment than truth!

Probability, utility and decisions

- The agent can use *probability theory* to reason about uncertainty
- The agent can use *utility theory* for rational selection of actions based on preferences
- *Decision theory* is a general theory for combining probability with rational decisions

Decision theory = Probability theory
+ Utility theory

Decision theoretic agent

function DT-AGENT(*percept*) **returns** an *action*

persistent: *belief-state*, probabilistic beliefs about
the state of the world

action, the agent's action

update *belief-state* based on *action* and *percept*

calculate outcome probabilities for actions, given
action descriptions and current *belief-state*

select *action* with highest expected utility given
probabilities of outcomes and utility information

return *action*

Basic probability notation

- A *probability model* is a set of *propositions*, expressed in terms of *random variables* with *domains*
  - Boolean – E.g. *Cavity*: <*true, false*>
  - Discrete – E.g. *Weather*: <*sunny, rainy, cloudy, snow*>
  - Continuous – E.g. *Index*: [0, 1]
- An *atomic event* is an assignment of particular values to all variables of the domain
  - E.g. $Cavity = false \wedge Toothache = true$
  - Mutually exclusive (only one event can be true at a time)
  - Exhaustive (at least one must be true)
- *Prior* (unconditional) probability of a proposition: $P(A)$
  - $P(Cavity) = 0.1$ … i.e. $P(Cavity = true) = 0.1$

Basic probability notation (cont.)

- Probability *distribution* of variable **P**(*v*)
  - **P**(*Weather*) = (0.7, 0.2, 0.08, 0.02)
- *Joint* probability distribution
  - Table of probabilities for all *combinations*: **P**(v_1, v_2)
  - **P**(*Weather*, *Cavity*) is a 4 x 2 table of probabilities (must sum to 1)
  - *Full joint distribution*: all domain variables included
- *Conditional* (posterior) probability: $P(A|B)$
  - $P(Cavity|Toothache) = 0.8$
- *Product rule:*
  - $P(A \wedge B) = P(A|B)\ P(B)$
  - $P(A \wedge B) = P(B|A)\ P(A)$
  - $P(A|B) = P(A \wedge B) / P(B)$

Axioms of probability

- Basic axioms

$$0 \leq P(A) \leq 1$$

$$P(True) = 1 \quad P(False) = 0$$

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$

- All other properties can be derived, e.g.

$$P(A \vee \neg A) = P(A) + P(\neg A) - P(A \wedge \neg A)$$

$$P(True) = P(A) + P(\neg A) - P(False)$$

$$1 = P(A) + P(\neg A)$$

$$P(\neg A) = 1 - P(A)$$

Inference using full joint distribution

	Toothache		¬ *Toothache*	
	Catch	¬ *Catch*	*Catch*	¬ *Catch*
Cavity	0.108	0.012	0.072	0.008
¬*Cavity*	0.016	0.064	0.144	0.576

Σ=1

- Probability of *combination* of events
  - $P(cavity \vee toothache)$ = 0.108+0.012+0.072+0.008+0.016+0.064 = 0.28
- *Unconditional* probability of a variable (marginalization)
  - $P(cavity)$ = 0.108+0.012+0.072+0.008 = 0.2
- *Conditional* probability (using product rule)
  - $P(cavity|toothache) = P(cavity \wedge toothache)/P(toothache)$
    = (0.108+0.012)/(0.108+0.012+0.16+0.064) = 0.6
- Problem: This approach does not scale up!
  - Table size and calculation time is $O(2^n)$ for *n* Boolean variables

Bayes' rule

- Bayes' rule: $P(B|A) = \frac{P(A|B)P(B)}{P(A)}$
- Easily derived from product rule
- Underlies most modern AI systems for probabilistic inference
- Main application:
  - How to use *prior* and *causal* knowledge (*cause* $\Rightarrow$ *effect*) to derive *diagnosis* (*effect* $\Rightarrow$ *cause*)

$$P(cause \mid effect) = \frac{P(effect \mid cause)P(cause)}{P(effect)}$$

- Use of causal knowledge is crucial in making probabilistic reasoning sufficiently robust in applications

Simple example of using Bayes' rule

- Diagnosing meningitis
  - Prior probability of meningitis: $P(M)=1/50000$
  - Prior probability of stiff neck: $P(S)=1/20$
  - Meningitis *causes* stiff neck: $P(S|M)=1/2$
- What is probability that a patient with stiff neck has meningitis?

$$P(M|S) = \frac{P(S|M)P(M)}{P(S)} = \frac{0.5 \times 1/50000}{1/20} = 1/5000$$

- Very low probability of meningitis (because prior prob. of stiff neck >> prior prob. of meningitis)

Combining evidence

- We can use Bayes' rule to find probability of a state given several pieces of *evidence*

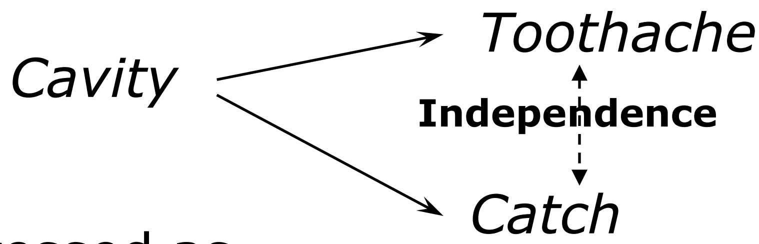
$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

$P(Cavity|Toothache \wedge Catch) =$
$P(Toothache \wedge Catch|Cavity)P(Cavity)/P(Toothache \wedge Catch)$

- For this to work we need conditional probability for all combinations of evidence variables
- In general case there is an *exponential* number of conditional probabilities
  - For *n* Boolean evidence variables, we need 2^n conditional probabilities
- This led AI researcher away from probability theory and towards more *ad hoc* systems

Conditional independence

- Using Bayes' rule is simplified in situations of *conditional independence* between variables



Toothache and *Catch* are conditionally independent given presence/absence of *Cavity*

- Expressed as

  $P(Toothache \wedge Catch \mid Cavity) =$
  $P(Toothache \mid Cavity)P(Catch \mid Cavity)$

- Simplifies evidence combination with Bayes' rule

  $P(Cavity \mid Toothache \wedge Catch) =$
  $P(Toothache \wedge Catch \mid Cavity)P(Cavity)/P(Toothache \wedge Catch) =$
  $P(Toothache \mid Cavity)P(Catch \mid Cavity)P(Cavity)/P(Toothache \wedge Catch)$

- Combining evidence does not need information for all combinations of evidence variables, only separate conditionals

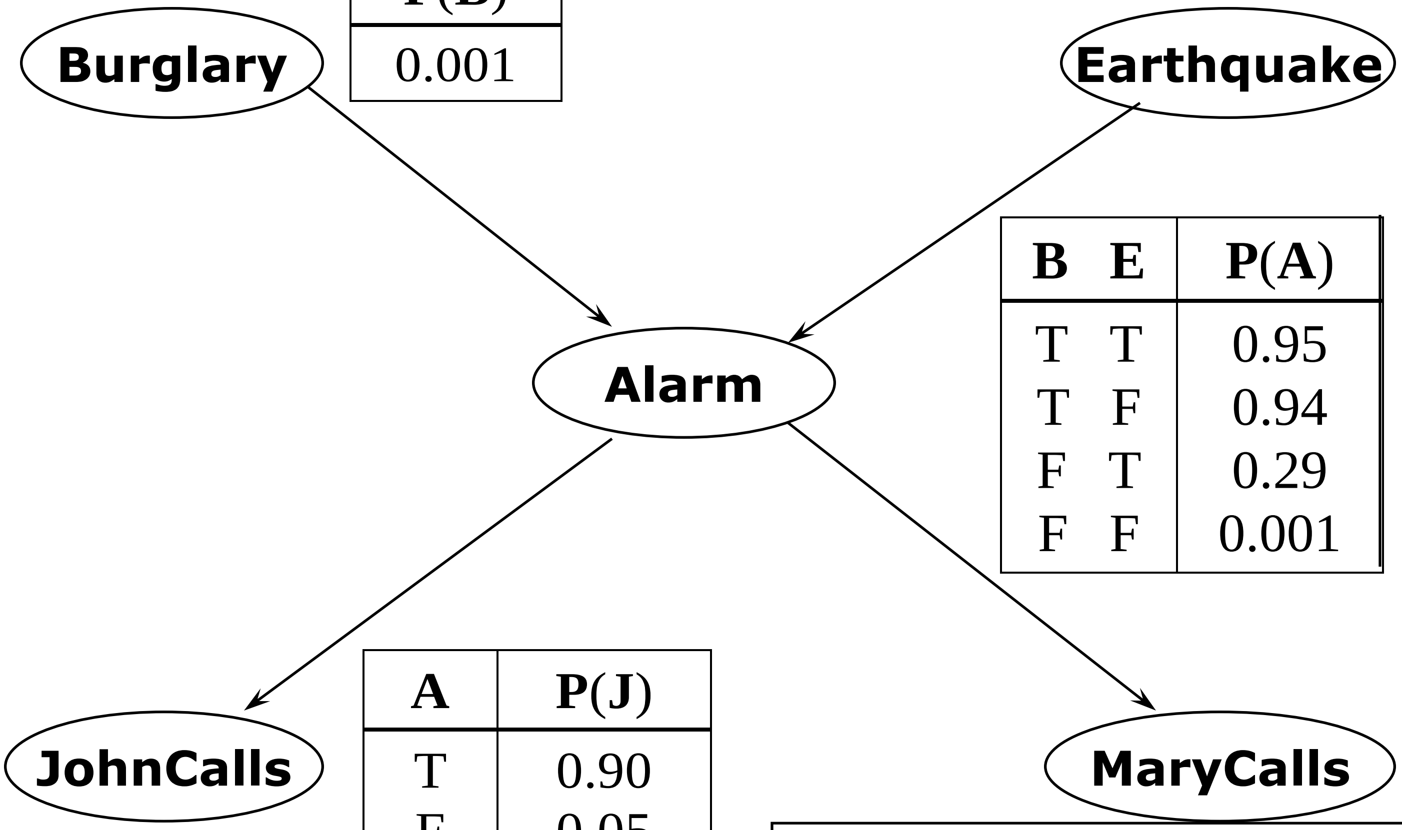
Bayesian networks

- *Bayesian networks* represent dependencies between variables and give concise specification of joint probability distribution
- A Bayesian network is a graph where
  - √ A set of *random variables* are the nodes of the graph
  - √ A set of *directed links* connects pairs of nodes
  - √ A link from X to Y means that X has *direct influence* on Y, or is the *parent* of Y
  - √ Each node has a *conditional probability table* (CPT) that quantifies effects parent nodes have on the node
  - √ The graph has *no directed cycles* (it is a directed, acyclic graph, or DAG)
- It can be shown that a Bayesian network is a representation of the *joint probability distribution*

Incremental construction of Bayesian network

- General procedure
  - Choose set of relevant domain variables X_i
  - Choose an ordering of the variables, preferably using causal domain knowledge (“root causes first”, etc.)
  - While there are variables left
    - Pick the next variable X_i and add node for it
    - Set *Parents*(X_i) to minimal set of nodes already in net such that X_i depends directly only on these nodes
    - Define conditional probability table for X_i
- Guarantees that
    - Network is acyclic
    - Axioms of probability are satisfied

Example Bayesian network



P(B)
0.001

P(E)
0.001

B	E	P(A)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

Each CPT also has e.g. ¬ P(A) row with prob. 1-P(A)

A	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

Q: Given who calls, what is the prob. of Burglary?

Inference in Bayesian networks

- Probabilistic inference procedure computes posterior probability distribution for a set of *query variables*, given exact values for some *evidence variables*

$$\mathbf{P}(Query|Evidence)$$

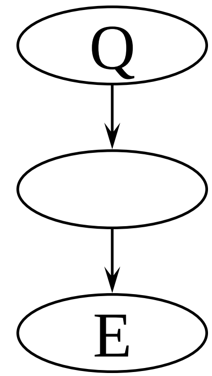
- An agent gets values for evidence variables from its percepts and queries for other variables in order to decide on action, using two functions
- *Exact* and *approximate* (Monte Carlo) methods have been developed for Bayesian inference

Example of Bayesian inference

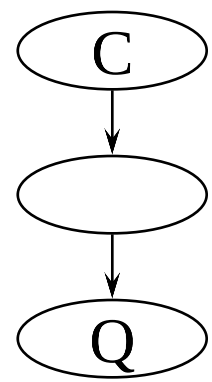
- What is $P(Burglary|JohnCalls)$? ("Diagnosis")
- Incorrect reasoning:
  - Since $P(JohnCalls|Alarm) = 0.9$, $P(Burglary|JohnCalls)$ "should be" 0.8-0.9
  - Incorrect due to false alarms: $P(JohnCalls|\neg Alarm) = 0.05$
- Correct reasoning
  - Using Bayes' rule $P(Burglary|JohnCalls) = 0.016$

Types of Bayesian inference

- Diagnostic inference
  - From effects to causes
  - E.g. given that *JohnCalls*, infer that $P(Burglary|JohnCalls) = 0.016$



- Causal inference
  - From causes to effects
  - E.g. from *Burglary*, infer that $P(JohnCalls|Burglary) = 0.86$ and $P(MaryCalls|Burglary) = 0.67$



Other uses of Bayesian networks

- *Make decisions* based on probabilities in the network and agent utilities
- Decide which *additional evidence* variables should be observed in order to gain useful information
- Perform *sensitivity analysis* to understand which aspects of model have greatest impact on probabilities of query variables
- *Explain results* of probabilistic inference to users

Other approaches to uncertainty

- Default reasoning
- Rule-based methods
- Dempster-Shafer theory
- Fuzzy sets and fuzzy logic

Default reasoning

- Human judgmental reasoning is "qualitative" and does not rely on numerical probabilities
- Default reasoning:
  - √ Assume default state until new evidence presents itself
  - √ If new evidence, conclusion may be retracted
- This kind of reasoning is called *nonmonotonic*, because set of beliefs may both grow and shrink. Problems:
  - √ Unclear semantic status of default rules
  - √ What if two matching default rules have contradictory conclusions
  - √ Managing retraction of beliefs (*truth maintenance* systems)
  - √ How to use default rules for making decisions
- No default reasoning system has solved all issues, and most systems are formally undecidable, and very slow in practice

Rule-based methods

- Logical and rule-based reasoning systems have useful properties that probabilistic systems lack
  - *Locality*: Can use A => B independent of other rules
  - *Detachment*: Can use belief independent of its justification
  - *Truth-functionality*: Truth of complex sentence follows from truth of components
- Attempts have been made to modify rule systems by attaching *degrees of belief* to propositions/rules
  - Best known is the *certainty factor* model (*Mycin*, ca. 1980)
- The problem is that above properties are *not appropriate* for uncertain reasoning, and rule-based approaches to uncertainty have fallen out of use

Dempster-Shafer theory

- *Dempster-Shafer theory* deals with the distinction between *uncertainty* and *ignorance*
- Instead of computing probability $P(X)$, it computes probability $Bel(X)$ that evidence supports proposition
- Example
  - For unknown possibly non-fair coin, $Bel(Heads)=Bel(\neg Heads)=0$
  - If 90% certain that coin is fair ($P(Heads)=0.5$), $Bel(Heads)=0.5\text{x}0.9=0.45$, $Bel(\neg Heads)=0.45$
- One interpretation of Dempster-Shafer is that it calculates a *probability interval*
  - *Heads* interval for unknown possibly non-fair coin: [0, 1]
  - Interval for 90% certain that coin is fair: [0.45, 0.55]
- Width of the interval helps decide when more evidence is required

Fuzzy sets and fuzzy logic

- *Fuzzy set theory* is a means of specifying how well an object satisfies a vague description
  - E.g. Is “Nate is tall” true if Nate is 5′ 10′′?
- In fuzzy set theory, *TallPerson* is a fuzzy predicate and the truth value of *TallPerson*(*X*) is in [0, 1]
- The fuzzy predicate defines a *fuzzy set* that does not have sharp boundaries, i.e. not uncertainty about the world, but uncertainty about the meaning of “tall”, and many claim that fuzzy set theory is not for uncertain reasoning at all
- *Fuzzy logic* can be used to determine truth value of complex sentence from truth of its components. However, fuzzy logic is inconsistent with first-order logic
- Despite this, fuzzy logic has been commercially successful, e.g. in automatic transmissions, trains, and video cameras

Summary

- *Uncertainty* arises because of both "laziness" and theoretical/practical ignorance, and cannot be avoided in complex worlds
- Basic probability include *prior probability* and *conditional probabilities*
- *Bayes' rule* allows unknown probabilities to be computed from known, stable ones
- Bayes' rule can be used to infer *diagnostic* conclusions from *causal* rules and prior probabilities
- In general case, combining many pieces of evidence requires many conditional probabilities

Summary (cont.)

- *Conditional independence* due to direct causal relationships allows efficient use of *Bayes* rule
- *Bayesian networks* are a natural way to represent conditional independence information, and is a complete representation of the joint probability distribution, but exponentially smaller
- Bayesian networks can reason *causally* or *diagnostically* (as well as in other ways)
- Many *alternative* uncertainty reasoning methods exist, incl. default reasoning, rule-based methods, Dempster-Shafer theory, fuzzy sets and fuzzy logic