Lecture 31 of 42

Reasoning under Uncertainty: Uncertain Inference Concluded
Discussion: Fuzzy Reasoning & D-S Theory

William H. Hsu
Department of Computing and Information Sciences, KSU

KSOL course page: http://snipurl.com/v9v3
Course web site: http://www.kddresearch.org/Courses/CIS730
Instructor home page: http://www.cis.ksu.edu/~bhsu

Reading for Next Class:
Review Chapters 13 – 14, R&N
Dempster-Shafer theory: http://en.wikipedia.org/wiki/Dempster-Shafer_theory
Fuzzy logic: http://en.wikipedia.org/wiki/Fuzzy_logic

Lecture Outline

- Reading for Next Class: Sections 14.1 – 14.2 (p. 492 – 499), R&N 2e
- Last Class: Uncertainty, Probability, 13 (p. 462 – 486), R&N 2e
  - Where uncertainty is encountered: reasoning, planning, learning (later)
  - Sources: sensor error, incomplete/inaccurate domain theory, randomness
- Today: Probability Intro, Continued, Chapter 13, R&N 2e
  - Why probability
    - Axiomatic basis: Kolmogorov
    - With utility theory: sound foundation of rational decision making
  - Joint probability
  - Independence
  - Probabilistic reasoning: inference by enumeration
  - Conditioning
    - Bayes's theorem (*aka* Bayes' rule)
    - Conditional independence
- Coming Week: More Applied Probability, Graphical Models

Acknowledgements

Stuart J. Russell
Professor of Computer Science
Chair, Department of Electrical Engineering and Computer Sciences
Smith-Zadeh Prof. in Engineering

Peter Norvig
Director of Research

Lotfali Asker-Zadeh (Lotfi A. Zadeh)
Professor of Computer Science
Department of Electrical Engineering and Computer Sciences
Director, Berkeley Initiative in Soft Computing
University of California - Berkeley

Russell, S. J.
University of California, Berkeley
http://www.eecs.berkeley.edu/~russell/

Norvig, P.
http://norvig.com/

Slides from:
http://aima.eecs.berkeley.edu

Zadeh, L. A.
University of California, Berkeley
http://bit.ly/39shSQ

Probability: Basic Definitions and Axioms

- Sample Space (Ω): Range of Random Variable *X*
- Probability Measure *Pr*(•)
  - Ω denotes range of observations; X: Ω
  - Probability *Pr*, or *P*: *measure* over power set 2^Ω - event space
  - In general sense, $Pr(X = x \in \Omega)$ is measure of belief in $X = x$
    - ⇨ $P(X = x) = 0$ or $P(X = x) = 1$: plain (*aka* categorical) beliefs
    - ⇨ Can't be revised; all other beliefs are subject to revision
- Kolmogorov Axioms
  - 1. $\forall x \in \Omega \,.\, 0 \le P(X = x) \le 1$
  - 2. $P(\Omega) \equiv \sum_{x \in \Omega} P(X = x) = 1$
  - 3. $\forall X_1, X_2, \ldots \ni i \neq j \Rightarrow X_i \wedge X_j = \varnothing$.

$$P\left(\bigcup_{i=1}^{\infty} X_i\right) = \sum_{i=1}^{\infty} P(X_i)$$

- Joint Probability: $P(X_1 \wedge X_2) \equiv$ Prob. of Joint Event $X_1 \wedge X_2$
- Independence: $P(X_1 \wedge X_2) = P(X_1) \bullet P(X_2)$

Evidential Reasoning – Inference by Enumeration Approach

Let $\mathbf{X}$ be all the variables. Typically, we want
the posterior joint distribution of the query variables $\mathbf{Y}$
given specific values $\mathbf{e}$ for the evidence variables $\mathbf{E}$

Let the hidden variables be $\mathbf{H} = \mathbf{X} - \mathbf{Y} - \mathbf{E}$

Then the required summation of joint entries is done by summing out the hidden variables:

$$\mathbf{P}(\mathbf{Y}|\mathbf{E}=\mathbf{e}) = \alpha\mathbf{P}(\mathbf{Y},\mathbf{E}=\mathbf{e}) = \alpha\Sigma_{\mathbf{h}}\mathbf{P}(\mathbf{Y},\mathbf{E}=\mathbf{e},\mathbf{H}=\mathbf{h})$$

The terms in the summation are joint entries because $\mathbf{Y}$, $\mathbf{E}$, and $\mathbf{H}$ together exhaust the set of random variables

Obvious problems:
1) Worst-case time complexity $O(d^n)$ where d is the largest arity
2) Space complexity $O(d^n)$ to store the joint distribution
3) How to find the numbers for $O(d^n)$ entries???

Bayes's Theorem (*aka* Bayes' Rule)

Product rule $P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$

$$\Rightarrow \text{ Bayes' rule } P(a|b) = \frac{P(b|a)P(a)}{P(b)}$$

or in distribution form

$$\mathbf{P}(Y|X) = \frac{\mathbf{P}(X|Y)\mathbf{P}(Y)}{\mathbf{P}(X)} = \alpha\mathbf{P}(X|Y)\mathbf{P}(Y)$$

Useful for assessing diagnostic probability from causal probability:

$$P(Cause|Effect) = \frac{P(Effect|Cause)P(Cause)}{P(Effect)}$$

E.g., let M be meningitis, S be stiff neck:

$$P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{0.8 \times 0.0001}{0.1} = 0.0008$$

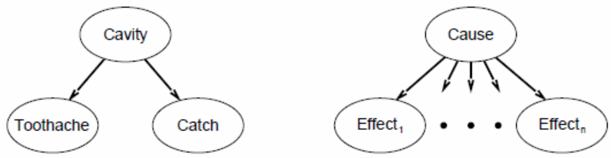
Note: posterior probability of meningitis still very small!

Bayes' Rule & Conditional Independence

$$\mathbf{P}(Cavity|toothache \wedge catch)$$
$$= \alpha\,\mathbf{P}(toothache \wedge catch|Cavity)\mathbf{P}(Cavity)$$
$$= \alpha\,\mathbf{P}(toothache|Cavity)\mathbf{P}(catch|Cavity)\mathbf{P}(Cavity)$$

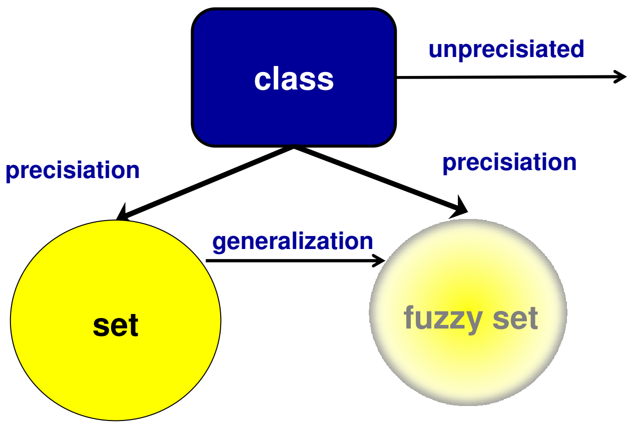
This is an example of a naive Bayes model:

$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause)\Pi_i\mathbf{P}(Effect_i|Cause)$$



Total number of parameters is linear in n

Non-Probabilistic Representation [1]: Concept of Fuzzy Set



Informally, a fuzzy set, *A*, in a universe of discourse, *U*, is a class with a fuzzy boundary.

Adapted from slide © 2008 L. A. Zadeh, UC Berkeley http://bit.ly/39shSQ

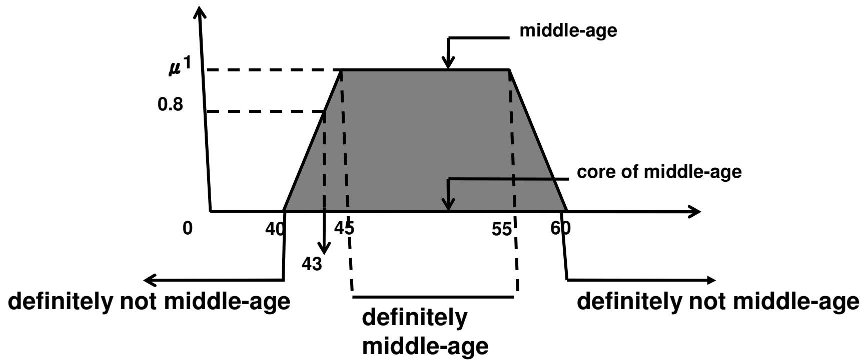
Non-Probabilistic Representation [2]: Precisiation & Degree of Membership

- Set *A* in *U*: Class with Crisp Boundary
- Precisiation: Association with Function whose Domain is *U*
- Precisiation of Crisp Sets
  - Through association with (Boolean-valued) characteristic function
  - $c_A: U \rightarrow \{0, 1\}$
- Precisiation of Fuzzy Sets
  - Through association with membership function
  - $\mu_A: U \rightarrow [0, 1]$
  - $\mu_A(u)$, $u \in U$, represents grade of membership of *u* in *A*
- Degree of Membership
  - Membership in *A*: matter of degree
  - "In fuzzy logic everything is or is allowed to be a matter of degree." – Zadeh

Adapted from slide © 2008 L. A. Zadeh, UC Berkeley http://bit.ly/39shSQ

Non-Probabilistic Representation [3]: Fuzzy Set Example – Middle-Age

- "Linguistic" Variables: Qualitative, Based on Descriptive Terms
- Imprecision of Meaning = Elasticity of Meaning
- Elasticity of Meaning = Fuzziness of Meaning



Adapted from slide © 2008 L. A. Zadeh, UC Berkeley http://bit.ly/39shSQ

Automated Reasoning using Probability: Inference Tasks

Simple queries: compute posterior marginal $\mathbf{P}(X_i|\mathbf{E}=\mathbf{e})$
e.g., $P(NoGas|Gauge=empty, Lights=on, Starts=false)$

Conjunctive queries: $\mathbf{P}(X_i, X_j|\mathbf{E}=\mathbf{e}) = \mathbf{P}(X_i|\mathbf{E}=\mathbf{e})\mathbf{P}(X_j|X_i, \mathbf{E}=\mathbf{e})$

Optimal decisions: decision networks include utility information;
probabilistic inference required for $P(outcome|action, evidence)$

Value of information: which evidence to seek next?

Sensitivity analysis: which probability values are most critical?

Explanation: why do I need a new starter motor?

Choosing Hypotheses

- **Bayes's Theorem**

$$P(h\,|\,D) = \frac{P(D\,|\,h)P(h)}{P(D)} = \frac{P(h \wedge D)}{P(D)}$$

- **MAP Hypothesis**
  - Generally want most probable hypothesis given training data
  - Define: $\arg\max_{x \in \Omega} [f(x)] \equiv$ value of x in sample space Ω with highest $f(x)$
  - Maximum *a posteriori* hypothesis, h_{MAP}

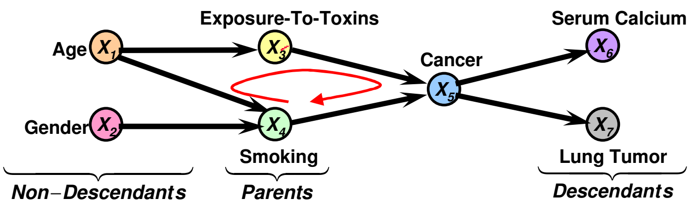
$$\begin{aligned} h_{MAP} &= \arg\max_{h \in H} P(h\,|\,D) \\ &= \arg\max_{h \in H} \frac{P(D\,|\,h)P(h)}{P(D)} \\ &= \arg\max_{h \in H} P(D\,|\,h)P(h) \end{aligned}$$

- **ML Hypothesis**
  - Assume that $p(h_i) = p(h_j)$ for all pairs i, j (uniform priors, i.e., $P_H \sim$ Uniform)
  - Can further simplify and choose maximum likelihood hypothesis, h_{ML}

$$h_{ML} = \arg\max_{h_i \in H} P(D\,|\,h_i)$$

Graphical Models of Probability

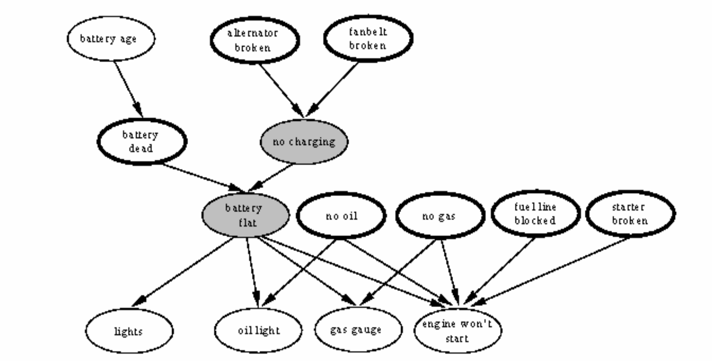
- **Conditional Independence**
  - *X* is conditionally independent (CI) from *Y* given *Z* iff $P(X \mid Y, Z) = P(X \mid Z)$ for all values of *X, Y*, and *Z*
  - Example: $P(Thunder \mid Rain, Lightning) = P(Thunder \mid Lightning) \Leftrightarrow T \perp R \mid L$
- **Bayesian (Belief) Network**
  - Acyclic directed graph model $B = (V, E, \Theta)$ representing *CI assertions* over Θ
  - Vertices (nodes) *V*: denote events (each a random variable)
  - Edges (arcs, links) *E*: denote conditional dependencies
- **Markov Condition for BBNs (Chain Rule):** $P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))$
- **Example BBN**



P(*20s*, *Female*, *Low*, *Non-Smoker*, *No-Cancer*, *Negative*, *Negative*)
= $P(T) \cdot P(F) \cdot P(L \mid T) \cdot P(N \mid T, F) \cdot P(N \mid L, N) \cdot P(N \mid N) \cdot P(N \mid N)$

Evidential Reasoning: Example – Car Diagnosis

Initial evidence: engine won't start
Testable variables (thin ovals), diagnosis variables (thick ovals)
Hidden variables (shaded) ensure sparse structure, reduce parameters



Adapted from slide © 2004 S. Russell & P. Norvig. Reused with permission.

Tools for Building Graphical Models

- **Commercial Tools: *Ergo, Netica, TETRAD, Hugin***
- ***Bayes Net Toolbox (BNT)* – Murphy (1997-present)**
  - **Distribution page** http://http.cs.berkeley.edu/~murphyk/Bayes/bnt.html
  - **Development group** http://groups.yahoo.com/group/BayesNetToolbox
- ***Bayesian Network tools in Java (BNJ)* – Hsu *et al.* (1999-present)**
  - **Distribution page** http://bnj.sourceforge.net
  - **Development group** http://groups.yahoo.com/group/bndev
  - **Current (re)implementation projects for KSU KDD Lab**
    - **Continuous state: Minka (2002) – Hsu, Guo, Li**
    - **Formats: XML BNIF (MSBN), Netica – Barber, Guo**
    - **Space-efficient DBN inference – Meyer**
    - **Bounded cutset conditioning – Chandak**

References: Graphical Models & Inference

- **Graphical Models**
  - **Bayesian (Belief) Networks tutorial – Murphy (2001)**
    http://www.cs.berkeley.edu/~murphyk/Bayes/bayes.html
  - **Learning Bayesian Networks – Heckerman (1996, 1999)**
    http://research.microsoft.com/~heckerman
- **Inference Algorithms**
  - **Junction Tree (Join Tree, L-S, *Hugin*): Lauritzen & Spiegelhalter (1988)**
    http://citeseer.nj.nec.com/huang94inference.html
  - **(Bounded) Loop Cutset Conditioning: Horvitz & Cooper (1989)**
    http://citeseer.nj.nec.com/shachter94global.html
  - **Variable Elimination (Bucket Elimination, *ElimBel*): Dechter (1986)**
    http://citeseer.nj.nec.com/dechter96bucket.html
  - **Recommended Books**
    - **Neapolitan (1990) – *out of print*; see Pearl (1988), Jensen (2001)**
    - **Castillo, Gutierrez, Hadi (1997)**
    - **Cowell, Dawid, Lauritzen, Spiegelhalter (1999)**
  - **Stochastic Approximation**
    http://citeseer.nj.nec.com/cheng00aisbn.html

Terminology

- **Uncertain Reasoning: Inference Task with Uncertain Premises, Rules**
- **Probabilistic Representation**
  - Views of probability
    - Subjectivist: measure of belief in sentences
    - Frequentist: likelihood ratio
    - Logicist: counting evidence
  - Founded on Kolmogorov axioms
    - Sum rule
    - Prior, joint *vs.* conditional
    - Bayes's theorem & product rule: P(A | B) = (P(B | A) * P(A)) / P(B)
  - Independence & conditional independence
- **Probabilistic Reasoning**
  - Inference by enumeration
  - Evidential reasoning

Summary Points

- **Last Class: Reasoning under Uncertainty and Probability (Ch. 13)**
  - Uncertainty is pervasive
  - What are we uncertain about?
- **Today: Chapter 13 Concluded, Preview of Chapter 14**
  - Why probability
    - Axiomatic basis: Kolmogorov
    - With utility theory: sound foundation of rational decision making
  - Joint probability
  - Independence
  - Probabilistic reasoning: inference by enumeration
  - Conditioning
    - Bayes's theorem (*aka* Bayes' rule)
    - Conditional independence
- **Coming Week: More Applied Probability**
  - Graphical models as KR for uncertainty: Bayesian networks, *etc.*
  - Some inference algorithms for Bayes nets