CS480 – Artificial Intelligence Fall 2015

Topic: Uncertainty
Chapter: 13
Date: 10/21

Mustafa Bilgic

http://www.cs.iit.edu/~mbilgic

Medical Diagnosis - Toothache

- Toothache $\Rightarrow$ Cavity
  - What's wrong?
    - Not all toothaches are due to cavity
- Toothache $\Rightarrow$ Cavity $\vee$ GumProblem $\vee$ Abscess $\vee$ …
  - What's wrong?
    - We have to add almost an unlimited number of possible problems
- Cavity $\Rightarrow$ Toothache
  - What's wrong?
    - Not all cavities cause toothache

FOL

- Fails for three main reasons

1. **Laziness**
   - Too much work to list all premises and conclusions
   - Too hard to use such rules
2. **Theoretical ignorance**
   - No complete theory for the domain
3. **Practical ignorance**
   - Even if we knew all the rules, some information might be missing; e.g., lab tests for diagnosis

Sources of uncertainty

- **Uncertainty in knowledge**
  - E.g., We do not know all the causes of all the diseases
- **Uncertainty in actions**; we cannot list all the pre-conditions of actions
  - E.g., To be able to fly a plane from SFO to JFK, it must not be broken, the weather conditions have to be appropriate, the pilot must not be sick, you need to have enough fuel, …
- **Uncertainty in sensors**
  - E.g., lightning conditions for a camera might not be enough

TASKS

- Representation
  - What is the formal and appropriate language to represent uncertainty?
- Inference
  - How can we infer uncertainty before or after we gather more information?
- Decision making
  - How can a rational agent act in an uncertain world?

Probability Model

- It's all about the state the world is in, i.e., a possible world
- A possible world is an assignment of truth values to the predicates
  - Logical assertions rule out some of the possible worlds
    - E.g., cavity $\Rightarrow$ toothache rules out the worlds where cavity=true $\wedge$ toothache= false
  - Probabilistic reasoning determines how probable the various worlds are
    - E.g., a world where cavity=true $\wedge$ toothache=true is more probable than the world where cavity=true $\wedge$ toothache= false
- The set of all possible worlds is called the sample space
- The possible worlds are *mutually exclusive* and *exhaustive*
  - *Mutually exclusive:* the world can be only in one state
  - *Exhaustive:* the world has to be in one of the states

Probability Model

- A **probability model** associates a numerical probability P(w) with each possible world w
  - P(w) sums to 1 over all possible worlds
- An **event** is the set of possible worlds where a given predicate is true
  - Roll two dice; the possible worlds are (1,1), (1,2), …, (6,6); 36 possible worlds
  - Predicate = two dice sum to 10
  - Event = {(4,6), (5,5), (6,4)}

Axioms of Probability

1. The probability P(α) of a proposition α is a real number between 0 and 1
2. P(true) = 1, P(false) = 0
3. $P(a \lor b) = P(a) + P(b) - P(a \land b)$

$P(\neg a)$

- $P(a \vee \neg a) = P(a) + P(\neg a) - P(a \wedge \neg a)$
- $P(true) = P(a) + P(\neg a) - P(false)$
- $1 = P(a) + P(\neg a) - 0$
- $P(\neg a) = 1 - P(a)$
- Intuitive explanation:
  - The probability of all possible worlds is 1
  - Either a or $\neg a$ holds in one world
  - The worlds that a holds and the worlds that $\neg a$ holds are mutually exclusive and exhaustive

Random Variable

- Like CSP
  - A factored representation of the world; random variables
  - Each variable has a domain
  - Probabilities over the domain values of a variable sum to 1
    - The possible worlds where a random variable takes a certain value are mutually exclusive and exhaustive (from the viewpoint of that variable)
- E.g.
  - D_1: {1, 2, 3, 4, 5, 6}

Joint Distribution

- We have n random variables, $V_1, V_2, \ldots, V_n$
- We are interested in the probability of a possible world, where
  - $V_1=v_1, V_2=v_2, \ldots, V_n = v_n$
- $P(V_1, V_2, \ldots, V_n)$ associates a probability for each possible world ≡ the **joint distribution**
- How many entries are there, if we assume the variables are all binary?
- How is this related to the truth tables in logic?

Toothache example

Toothache	Cavity	P(T,C)
toothache	cavity	0.15
toothache	¬cavity	0.10
¬toothache	cavity	0.05
¬toothache	¬cavity	0.70

These probabilities are different from what is given in the textbook

Notation

- An upper case letter A_1 represents a variable and its all possible values
- A lower case letter a_1 represents a particular value of the variable A_1
- $P(A_1)$ represents a table/function specifying a probability for each possible value of A_1
- $P(A_1=a_1)$ represents a scalar value specifying the probability of $A_1=a_1$
- We often abbreviate $P(A_1=a_1)$ as $P(a_1)$

Prior and Posterior

- Prior probability
  - Probability of a proposition in the absence of any other information
  - E.g., $P(V_1, V_3, V_5)$
- Conditional/posterior probability
  - Probability of a proposition given another piece of information
  - E.g., $P(V_2, V_3 \mid V_5 = T, V_7 = F)$
  - $P(A \mid B) = P(A \wedge B) / P(B)$

Number of Parameters

- Assuming everything is binary
- $P(V_1)$ requires
  - 1 independent parameter
- $P(V_1, V_2, \ldots, V_n)$ requires
  - 2^n-1 independent parameters
- $P(V_1 | V_2)$ requires
  - 2 independent parameters
- $P(V_1, V_2, \ldots, V_n \mid V_{n+1}, V_{n+2}, \ldots, V_{n+m})$ requires
  - $2^m \times (2^n-1)$ independent parameters

Tasks

1. Representation
   - For now, assume the representation is a table for all possible entries; we'll revisit this idea later to come up with a more efficient representation
2. Inference
   - Let's see how we can use the joint distribution to answer various questions
3. Decision making

Marginalization

- Given $P(V_1, V_2, \ldots, V_n \mid V_{n+1}, V_{n+2}, \ldots, V_{n+m})$, where $n>0$ and $m \geq 0$, we can find, for example
  - $P(V_i, V_j, V_k \mid V_{n+1}, V_{n+2}, \ldots, V_{n+m})$ where $i, j, k < n$ by summing out all the irrelevant variables
- Examples

$$P(A) = \sum_B P(A,B) = P(A,B=r) + P(A,B=s) + P(A,B=b)$$

$$P(A,B|C)$$

$$P(A|C)$$

$$\sum_B P(A,B|C)$$

$P(C) = \sum_{T} P(T,C)$

$P(T,C)$

$P(C)$

LET'S ANSWER A FEW QUERIES

Toothache	Cavity	P(T,C)
toothache	cavity	0.15
toothache	¬cavity	0.10
¬toothache	cavity	0.05
¬toothache	¬cavity	0.70

C

c :

¬c :

- P(cavity) = ? $0.15 + 0.05 = 0.20$
- P(¬cavity) = ? $0.10 + 0.70 = 0.80$
- P(toothache) = ? $0.15 + 0.10 = 0.25$
- P(¬toothache) = ? $0.05 + 0.70 = 0.75$

Let's Answer a Few Queries

Toothache	Cavity	P(T,C)
toothache	cavity	0.15
toothache	¬cavity	0.10
¬toothache	cavity	0.05
¬toothache	¬cavity	0.70

- $P(\text{cavity} \mid \text{toothache}) = ?$
- $P(\text{cavity} \mid \neg\text{toothache}) = ?$
- $P(\neg\text{cavity} \mid \text{toothache}) = ?$
- $P(\neg\text{cavity} \mid \neg\text{toothache}) = ?$
- $P(\text{toothache} \mid \text{cavity}) = ?$
- $P(\neg\text{toothache} \mid \text{cavity}) = ?$
- $P(\text{toothache} \mid \neg\text{cavity}) = ?$
- $P(\neg\text{toothache} \mid \neg\text{cavity}) = ?$

BAYES' RULE

- P(B | A) = P(A | B)*P(B) / P(A)
- Example use
  - P(cause | effect) = P(effect | cause)*P(cause) / P(effect)
- Why is this useful?
  - Because in practice it is easier to get probabilities for P(effect|cause) and P(cause) than for P(cause|effect)
    - E.g., P(disease|symptoms) = P(symptoms|disease)*P(disease) / P(symptoms)
    - It is easier to know what symptoms diseases cause. It is harder to diagnose a disease given symptoms

Tuberculosis revisited

- Tuberculosis test
  - The test is 90% accurate
    - If you have TB, the test is positive with 90% probability
    - If you don't have TB, the test is negative with 90% probability
  - John takes the test and the result is positive
  - What is the probability that John has TB?
- Formally
  - $P(+ \mid TB) = 0.9$
  - $P(- \mid \neg TB) = 0.9$
  - $P(TB \mid +) = ?$

Tasks

1. Representation
   - For now, assume the representation is the joint distribution; we'll come back to it later to come up with a more tractable representation
2. Inference
   - Let's see how we can use the joint distribution to answer various questions
3. Decision making
   - Assuming we have an efficient way of representing the probability distribution, and an efficient way of answering probabilistic queries, how can we make decisions?

DECISION MAKING

- Instead of T/F, we have probabilities
- Preferences for certain outcomes (world states)
  - Utility theory
- Maximize expected outcome
  - Decision theory = probability theory + utility theory
- **Maximum Expected Utility** principle
  - An agent is rational if and only if it chooses the action that yields the highest expected utility, averaged over all the possible outcomes of the action

What's Next

- Efficient Representation and Inference
  - Chapter 14
- Making simple decisions / utility theory
  - Chapter 16