innovate achieve lead

BITS Pilani

Machine Learning(IS ZC464)

Session 3: Uncertainty Handling in Real World using Probability Theory

Uncertainty in real world

- Uncertainty in reaching New Delhi Airport in 5 hours from Pilani
  - Cab engine may or may not work at any moment
  - The route is diverted due to a procession on the way
  - The road condition is bad unexpectedly
  - The tire needs replacement

  Etc.
- A person having stomach ache can be told that he is suffering from ulcer, while in actual it may be gastritis or overeating

example

If

A person has pneumonia

Then

has fever

is pale

has cough

white blood cells count is low

- Certainty exists in obtaining symptoms if the disease is confirmed

Disease → symptoms

- For converse, it is uncertain that if a person has fever, and has cough, then has pneumonia, but if all symptoms are known then the disease can be inferred

$Fever(p) \land pale(p) \land cough(p) \land WBC(p) \rightarrow pneumonia(p)$

Types of uncertainty

- Disease → symptoms

  pneumonia may have other symptoms too

- Symptoms → disease

  these symptoms may be common in other diseases as well,

[but if **<u>all possible</u>** symptoms can be observed and are same for all patients, then more definiteness can be inroduced]

Real world scenario

- It is impossible to list all relevant components of the real world
- Many of the components behave with some uncertainty
- Due to system hardware limitation, representing all components of any real world situation may not be possible

Class assignment

- Analyze the weather on a day
  - It is cloudy (How much?)
    - Depends on individual belief
    - Belief can be based on experience
    - Experience may count on favorable situations
  - The day is humid
    - Is it the sufficient humidity that may cause rains
  - Is it certain that the clouds will rain.
    - The clouds may rain if certain other parameters are favourable.

Conventional reasoning

- Based on three assumptions
  - Predicate descriptions must be sufficient with respect to the application domain
  - The information base is consistent.
  - Through the inference rules, the known information grows monotonically
- Conventional methods follow closed world assumptions

Closed World Assumptions

- The closed world assumptions are based on the minimal model of the world.
- Any predicate not existing is false.

  Example: whether two cities are connected by a plane fight.

  - Check the list, if there is no direct flight, then we may infer that the cities are not connected.
- Exactly those predicates that are necessary for a solution are created.
- The closed world assumption affects the semantics of negation in reasoning.

Example : conventional reasoning

- Human(p) → mammal(p) ∧ intelligent(p) ∧kind(p) ∧ legs(p) ∧ eyes(p) ∧ ....

- mammal(John) ∧ legs(John) ∧ kind(John) ∧ eyes(John)
- What can be said about John's intelligence?
- Is John Intelligent?
- Is he not?
- Does lack of knowledge mean whether we are not sure that John is intelligent or we are sure that John is not intelligent

Uncertainty in First Order Logic

- $\forall$ x Bird(x) → Fly(x)
- Penguin is a bird. Does it fly?
- The above rule does not hold good for all birds (minimal world assumption)
- How can we generalize the rule?
- There can be a large number of predicates that can be constructed to represent a larger world
- $\forall$ x (Bird(x) $\wedge$ $\neg$Abnormal(x) → Fly(x))
- Uncertainty lies in the predicate abnormal.

Conventional reasoning

- Conventional logic is monotonic
- A set of predicates constitutes the knowledge base.
- The size of the KB keeps increasing if a new knowledge is added
- Pure methods of reasoning cannot handle KB with incomplete or uncertain knowledge

Nonmonotonic reasoning systems

- Addresses the problem of changing beliefs.
- Makes most reasonable assumptions in light of uncertain information.

Handling uncertain information using probability theory

- Probability theory deals with the degree of belief.
- Assigns numerical degree of belief between 0 and 1
- Handles the uncertainty that comes from laziness and ignorance
- The belief could be derived from
  - Statistical data
  - General rules
  - Combination of evidence sources

Example:

- $\forall$ p Symptom(p, toothache) $\Rightarrow$ Disease(p, cavity)
- The above for example can be said to carry a belief that 8 out of 10 patients have cavity when they had toothache.
- The probability associated with the above is 0.8.
- The belief may change if some more patients reach with pain and have different diseases.
- A probability of 0.8 does not mean that it is “80% true” but it is 80% degree of belief
- Degree of belief is different from degree of truth

Evidences

- The probability that a "patient has a cavity" is 0.8, depends on the agent's belief and not on the world.
- These beliefs depend on the percepts the agent has received so far
- These percepts constitute the evidence on which probability assertions are based.
- As new evidences add on, the probability changes.
- This is known as conditional probability.

Representing uncertain knowledge using probability

- Probability theory uses a language that is more expressive than the propositional logic
- The basic element of the language is the random variable.
- This random variable represents the real world whose status is initially known.
- The proposition asserts that a random variable has a particular value drawn from its domain

Types of random variables

- Boolean
  - domain is {true, false}
  - Example :
    - Cavity = true
- Discrete
  - domain is any set of integer values
  - Example
    - From domain { sunny, cloudy, rainy, snow} the variable may take whether = snow
- Continuous
  - domain takes values from real numbers

Atomic Events

- An atomic event is the complete specification of the state of the real world about which the agent is uncertain.
- Example:
  - Let the boolean random variables cavity and toothache constitute the real world then there are 4 atomic events

i. (Cavity = true) $\wedge$ (toothache= true)

ii. (Cavity = true) $\wedge$ (toothache= false)

iii. (Cavity = false) $\wedge$ (toothache= true)

iv. (Cavity = false) $\wedge$ (toothache= false)

Atomic events

- Mutually exclusive
- Set of all possible events is exhaustive (disjunction is true)
- Any proposition is logically equivalent to the disjunction of all atomic events that entail the truth of the proposition

Prior probability

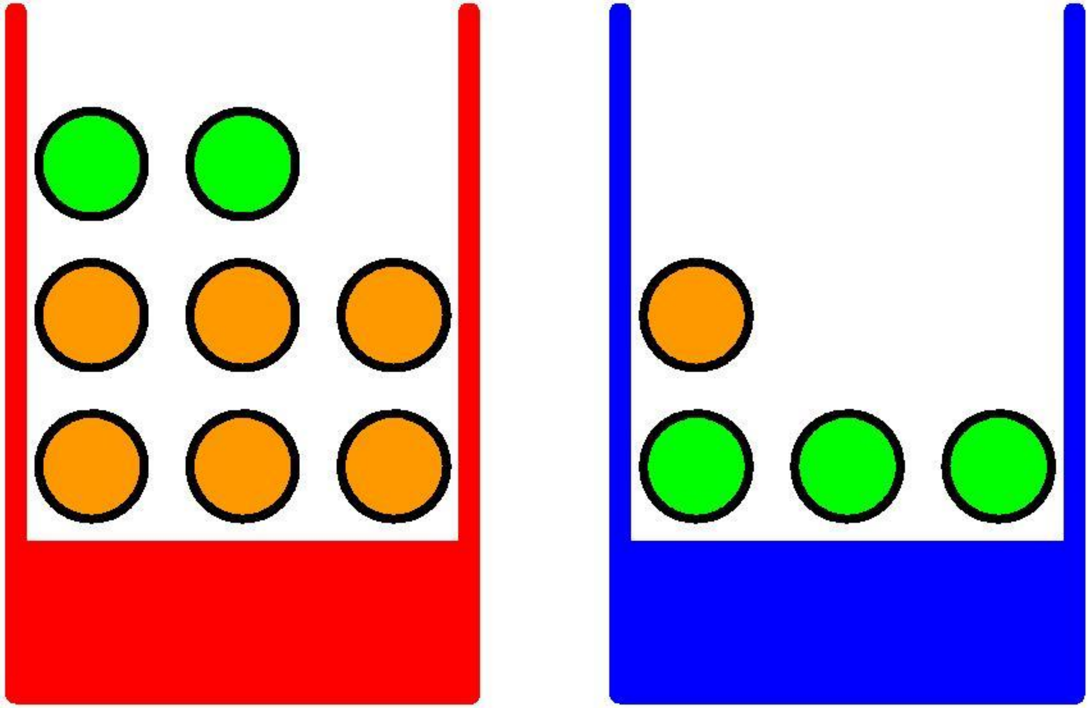
- The prior probability associated with proposition is the degree of belief in absence of any other information
- Example
  - P(cavity = true) = 0.1
  - P(cavity) = 0.1 [this is estimated based on the available information]
  - As some more information is available, the concept of conditional probability will be used to determine the P value

Computing Probability

- A bag contains 8 balls of which 6 are orange and 2 are green.
- A ball is chosen randomly from the bag.
- What is the probability that the ball is of green color? Answer = 2/8
- What is the probability that the ball is of orange color? Answer = 6/8

Probability Theory

Apples and Oranges kept in two bags of different colors

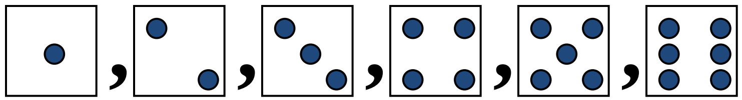


Computing Probability

- If a ball is to be chosen randomly from a bag, and a bag is chosen randomly, then how likely it is to select red bag? – Computed through experiments and multiple trials or is known apriori
- What is the probability that the ball selected from red bag is of green color?
- What is the probability that the ball selected from blue bag is of orange color?

Examples

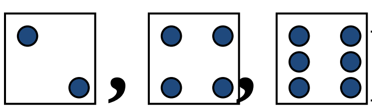
1. Rolling a die – outcomes

$S = \{$  $\}$

$= \{1, 2, 3, 4, 5, 6\}$

E = the event that an even number is rolled

$= \{2, 4, 6\}$

$= \{$  $\}$

Joint Probability

- This finds out how likely it is for two or more events to happen at the same time.
- Example
  - A patient has both cavity and toothache.
  - The joint probability is represented as

    $P(cavity \wedge toothache)$ or

    $P(cavity, toothache)$

Prior Probability Distribution

- Assume a discrete variable 'weather'
  - P(weather = sunny) = 0.4
  - P(weather = rainy) = 0.1
  - P(weather = cloudy) = 0.1
  - P(weather = snow) = 0.2
- The distribution is
  - P(weather)= { 0.4, 0.1, 0.1, 0.2}

Joint probability distribution

- P(weather, cavity) has 4x2 (=8) atomic events
- P(cavity, toothache, weather) has 2x2x4 (=16)
- Any probabilistic query can be answered using joint probability

Conditional Probability

- The intelligent agent may get new information about the random variables that make the domain
- The probabilities are recomputed
- Example
  - A bag/urn has 12 red colored balls and 8 blue balls.
  - The first trial, the probability of getting a red ball = 12/20
  - Second trial, the probability of getting red ball = 11/19

Conditional Probability

- Represented as $P(a|b)$
- $P(a|b) = P(a \wedge b) \;/\; P(b)$ for $P(b)>0$
- Also

$$P(a \wedge b) = P(a|b)\,P(b)$$

(Product Rule)

Axioms of Probability (Kolmogorov's Axioms)

- For any proposition a
  - $0 <= P(a) <= 1$
- True propositions have probability 1 and false propositions have value 0
  - $P(true)=1$, $P(false)=0$
- $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$

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

- Proof

$$a \wedge \neg a = \text{false}$$

$$a \vee \neg a = \text{true}$$

Using the third axiom of probability

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

$$\Longrightarrow P(\text{true}) = P(a) + P(\neg a) - P(\text{false})$$

$$\Longrightarrow P(\neg a) = 1 - P(a)$$

Proposition

- The probability of a proposition is equal to the sum of the probabilities of the atomic events in which it holds.
  - $P(a) = \Sigma P(ei)$ over all atomic events

Inference using Full Joint Distributions

- Joint distribution constructs the complete knowledge base
- Example
  - Let there be 2 random boolean variables representing the real world, say they are cavity and toothache

	toothache	**¬toothache**
Cavity	0.25	0.15
¬Cavity	0.10	0.50

P(cavity)= 0.25 + 0.15 = 0.4
P(toothache) = 0.25 + 0.10 = 0.35

Marginal Probability

- $P(Y) = \sum P(Y,z)$ (sum over all joint probabilities of Y with z)
  [Marginalization Rule]
- P(Y) is the distribution over Y obtained by summing out all the other variables from any joint distribution containing Y.
- Example:
  - P(cavity) = P(cavity, toothache) + P(cavity, $\neg$ toothache)
    = 0.25 + 0.15
    = 0.4

Conditioning

- $P(Y) = \sum P(Y,z)$

  $= \sum P(Y|z)\, P(z)$ (using product rule)

- Marginalization and Conditioning are useful rules for handling probability expressions

Computing conditional probabilities (only 2 random variables)

- P(Cavity | Toothache)

  $= P(cavity \wedge toothache) / P(toothache)$

  $= 0.25 / 0.35 = 0.7142$

- $P(\neg Cavity \mid toothache)$

  $= P(\neg Cavity \wedge Toothache) / P(toothache)$

  $= 0.1 / 0.35 = 0.2857$

Normalization Constant

- Normalization constant ensures that the conditional probabilities of events add up to 1.
- Example
  - P(cavity | toothache) = 0.999999 = 1
- Let α denote the normalization constant
  - Then the conditional probability

    $P(a|b) = P(a \wedge b) \;/\; P(b) \quad \text{for } P(b)>0$

  Becomes

    $P(a|b) = \alpha\, P(a \wedge b)$

Inference using Full Joint Distributions

- Let there be 3 random boolean variables representing the real world, say they are cavity, toothache and catch.
- We may still represent the joint probabilities as a table, shown below, but if we have more random variables, we simply use the propositions and their probabilities

Probability expressions

- $P(cavity, toothache, catch) = 0.06$
- $P(cavity, toothache, \neg catch) = 0.19$
- $P(cavity, \neg toothache, catch) = 0.05$
- $P(cavity, \neg toothache, \neg catch) = 0.10$
- $P(\neg cavity, toothache, catch) = 0.09$
- $P(\neg cavity, toothache, \neg catch) = 0.01$
- $P(\neg cavity, \neg toothache, catch) = 0.22$
- $P(\neg cavity, \neg toothache, \neg catch) = 0.28$

Compute P(cavity)
P(cavity,toothache)
P(toothache)

	toothache		**¬toothache**	
	Catch	¬catch	Catch	¬catch
Cavity	0.06	0.19	0.05	0.10
¬Cavity	0.09	0.01	0.22	0.28

Advantage of normalization constant

- Can help in generalizing the inference procedure
- $P(X \mid e) = \alpha\, P(X,e)$

$$= \alpha\, (P(X,e\,,y1) + P(X, e, y2))$$

Example: $P(cavity \mid toothache) = P(cavity, toothache)/P(toothache)$

$= (P(cavity,toothache,catch) + P(cavity,toothache, \neg catch))/ P(toothache)$

Probabilistic queries using joint probability distribution

- These queries are answered using the joint probability distribution.
- The joint probability distribution is the knowledge base for inference using uncertain real world
- With n random variables, the size of the table becomes 2^n.
- Time to answer a query = $O(2^n)$
- When n is large, the method becomes almost impractical to work with.