Discussion Section 10

Feng Shi

University of California

shi.feng@cs.ucla.edu

June 8, 2018

Overview

1. Solution to First-Order Logic of last discussion
   - First Order Logic - Knowledge Engineering
   - First Order Logic - Inference
2. Bayesian Networks
   - Basic Probability Theory
   - Bayesian Networks
3. Decision Trees
   - Decision Tree Algorithms

First Order Logic - Knowledge Engineering

This exercise uses the function *MapColor* and predicates $In(x, y)$, $Borders(x, y)$, and $Country(x)$, whose arguments are geographical regions, along with constant symbols for various regions. In each of the following we give an English sentence and a number of candidate logical expressions. For each of the logical expressions, state whether it

1. correctly expresses the English sentence;
2. is syntactically invalid and therefore meaningless; or
3. is syntactically valid but does not express the meaning of the English sentence.

First Order Logic - Exercise (1)

1. Paris and Marseilles are both in France
   - $In(Paris \wedge Marseilles, France)$
   - $In(Paris, France) \wedge In(Marseilles, France)$
   - $In(Paris, France) \vee In(Marseilles, France)$

First Order Logic - Solution (1)

1. Paris and Marseilles are both in France
   - $In(Paris \wedge Marseilles, France)$
     Syntactically **invalid**. Cannot use conjunction inside a term;
   - $In(Paris, France) \wedge In(Marseilles, France)$
     Correct;
   - $In(Paris, France) \vee In(Marseilles, France)$
     Incorrect. Disjunction does not express "both".

First Order Logic - Exercise (2)

1. There is a country that borders both Iraq and Pakistan.
   - $\exists c \quad Country(c) \land Border(c, Iraq) \land Border(c, Pakistan)$
   - $\exists c \quad Country(c) \Rightarrow [Border(c, Iraq) \land Border(c, Pakistan)]$
   - $[\exists c \quad Country(c)] \Rightarrow [Border(c, Iraq) \land Border(c, Pakistan)]$
   - $\exists c \quad Border(Country(c), Iraq \land Pakistan)$

First Order Logic - Solution (2)

1. There is a country that borders both Iraq and Pakistan.
   - $\exists c \quad Country(c) \wedge Border(c, Iraq) \wedge Border(c, Pakistan)$
     Correct;
   - $\exists c \quad Country(c) \Rightarrow [Border(c, Iraq) \wedge Border(c, Pakistan)]$
     Incorrect, Use of implication in existential;
   - $[\exists c \quad Country(c)] \Rightarrow [Border(c, Iraq) \wedge Border(c, Pakistan)]$
     Syntactically **invalid**. Variable c used outside the scope of its quantifier;
   - $\exists c \quad Border(Country(c), Iraq \wedge Pakistan)$
     Syntactically **invalid**. Cannot use conjunction inside a term.

First Order Logic Inference - Exercise (1)

Suppose you are given the following axioms:

1. $0 \leq 3$
2. $7 \leq 9$
3. $\forall x \quad x \leq x$
4. $\forall x \quad x \leq x + 0$
5. $\forall x \quad x + 0 \leq x$
6. $\forall x, y \quad x + y \leq y + x$
7. $\forall w, x, y, z \quad w \leq y \wedge x \leq z \Rightarrow w + x \leq y + z$
8. $\forall x, y, z \quad x \leq y \wedge y \leq z \Rightarrow x \leq z$

(a) Give a **backward-chaining** proof of the sentence $7 \leq 3 + 9$.

(b) Give a **forward-chaining** proof of the sentence $7 \leq 3 + 9$.

First Order Logic Inference - Solution (1-a)

Goal G0: $7 \leq 3 + 9$ Resolve with (8) $\{x_1/7, z_1/3 + 9\}$.

Goal G1: $7 \leq y_1$ Resolve with (4) $\{x_2/7, y_1/7 + 0\}$. Succeeds.

Goal G2: $7 + 0 \leq 3 + 9$ Resolve with (8) $\{x_3/7 + 0, z_3/3 + 9\}$

Goal G3: $7 + 0 \leq y_3$ Resolve with (6) $\{x_4/7, y_4/0, y_3/0 + 7\}$

Succeeds.

Goal G4: $0 + 7 \leq 3 + 9$ Resolve with (7) $\{w_5/0, x_5/7, y_5/3, z_5/9\}$.

Goal G5: $0 \leq 3$. Resolve with (1). Succeeds.

Goal G6: $7 \leq 9$. Resolve with (2). Succeeds.

G4 succeeds.

G2 succeeds.

G0 succeeds.

First Order Logic Inference - Solution (1-b)

From (1), (2), (7) $\{w/0, x/7, y/3, z/9\}$ infer (9) $0 + 7 \leq 3 + 9$.
From (9), (6), (8) $\{x_1/0, y_1/7, x_2/0 + 7, y_2/7 + 0, z_2/3 + 9\}$ infer (10) $7 + 0 \leq 3 + 9$.
(x_1, y_1 are renamed variables in (6). x_2, y_2, z_2 are renamed variables in (8).)
From (4), (10), (8) $\{x_3/7, x_4/7, y_4/7 + 0, z_4/3 + 9\}$ infer (11) $7 \leq 3 + 9$.
(x_3 is a renamed variable in (4). x_4, y_4, z_4 are renamed variables in (8).)

Basic Probability Theory

Basic Probability Theory - Exercise (1)

Given the full joint distribution shown in table below,

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

Table: A full joint distribution for *Toothache*, *Cavity*, *Catch* world

Calculate the following:

1. $P(Toothache)$
2. $P(Cavity)$
3. $P(Toothache|Cavity)$
4. $P(Cavity|Toothache \vee Catch)$

Basic Probability Theory - Solution (1)

We can calculate the probabilities by addition:

1. $P(Toothache) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$
2. $P(Cavity) = \langle 0.2, 0.8 \rangle$
   (with $P(+Cavity) = 0.108 + 0.012 + 0.072 + 0.008 = 0.2$)
3. $P(Toothache|Cavity) = \langle (.108 + .012)/0.2, (0.072 + 0.008)/0.2 \rangle = \langle 0.6, 0.4 \rangle$
4. First compute $P(Toothache \vee Catch) = 0.108 + 0.012 + 0.016 + 0.064 + 0.072 + 0.144 = 0.416$, then

$$P(Cavity|Toothache \vee Catch)$$

$$\langle (0.108 + 0.012 + 0.072)/0.416, (0.016 + 0.064 + 0.144)/0.416 \rangle$$

$$= \langle 0.4615, 0.5384 \rangle$$

Consider two medical tests, A and B, for a virus. Test A is 95% effective at recognizing the virus when it is present, but has a 10% false positive rate (indicating that the virus is present, when it is not). Test B is 90% effective at recognizing the virus, but has a 5% false positive rate. The two tests use independent methods of identifying the virus. The virus is carried by 1% of all people. Say that a person is tested for the virus using only one of the tests, and that test comes back positive for carrying the virus. Which test returning positive is more indicative of someone really carrying the virus? Justify your answer mathematically.

Basic Probability Theory - Solution (2) I

Let V be the statement that the patient has the virus, and A and B the statements that the medical tests A and B returned positive, respectively. The problem statement gives:

$$P(V) = 0.01$$

$$P(A|V) = 0.95$$

$$P(A|\neg V) = 0.10$$

$$P(B|V) = 0.90$$

$$P(B|\neg V) = 0.05$$

The test whose positive result is more indicative of the virus being present is the one whose posterior probability, $P(V|A)$ or $P(V|B)$ is largest. One can compute these probabilities directly from the information given, finding that $P(V|A) = 0.0876$ and $P(V|B) = 0.1538$, so B is more indicative.

Basic Probability Theory - Solution (2) II

Equivalently, the questions is asking which test has the highest posterior odds ratio $P(V|A)/P(\neg V|A)$. From the odd form of Bayes theorem:

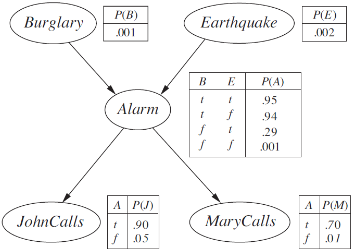
$$\frac{P(V|A)}{P(\neg V|A)} = \frac{P(A|V)}{P(A|\neg V)} \frac{P(V)}{P(\neg V)}$$

We see that the ordering is independent of the probability of V , and that we just need to compare the likelihood ratios $P(A|V)/P(A|\neg V) = 9.5$ and $P(B|V)/P(V|\neg V) = 18$ to find the answer.

Bayesian Networks

Bayesian Networks - Exercise (1)

Consider the Bayesian Network in figure



1. If no evidence is observed, are *Burglary* and *Earthquake* independent? Prove this from the numerical semantics and from the topological semantics.
2. If we observe $Alarm = true$, are *Burglary* and *Earthquake* independent? Justify your answer by calculating whether the probabilities involved satisfy the definition of conditional independence.

1. Yes. Numerically one can compute that $P(B, E) = P(B)P(E)$. Topologically B and E are d-separated by A.
2. We check whether $P(B, E|a) = P(B|a)P(E|a)$. First computing $P(B, E|a)$

$$P(B, E|a) = \alpha(a|B, E)P(B, E)$$

$$= \alpha \begin{cases} .95 \times 0.001 \times 0.002 & if\ B = b\ and\ E = e \\ .94 \times 0.001 \times 0.998 & if\ B = b\ and\ E = \neg e \\ .29 \times 0.999 \times 0.002 & if\ B = \neg b\ and\ E = e \\ .001 \times 0.999 \times 0.998 & if\ B = \neg b\ and\ E = \neg e \end{cases}$$

$$= \alpha \begin{cases} 0.0008 & if\ B = b\ and\ E = e \\ 0.3728 & if\ B = b\ and\ E = \neg e \\ 0.2303 & if\ B = \neg b\ and\ E = e \\ 0.3962 & if\ B = \neg b\ and\ E = \neg e \end{cases}$$

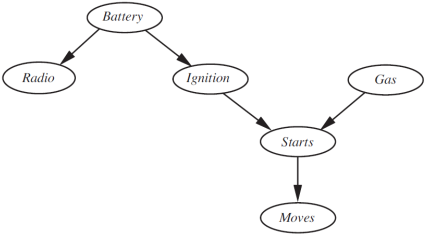
where α is a normalization constant. Checking whether $P(b, e|a) = P(b|a)P(e|a)$ we find

$$P(b, e|a) = 0.0008 \neq 0.0863 = 0.3736 \times 0.2311 = P(b|a)P(e|a)$$

showing that B and E are not conditionally independent given A.

Bayesian Networks - Exercise (2) I

Consider the network for car diagnosis shown in figure



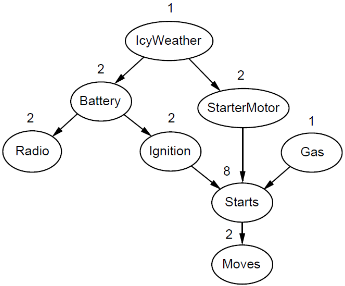
1. Extend the network with the Boolean variables *IcyWeather* and *StarterMotor*.
2. Give reasonable conditional probability tables for all the nodes.

Bayesian Networks - Exercise (2) II

3. How many independent values are contained in the joint probability distribution for eight Boolean nodes, assuming that no conditional independence relations are known to hold among them?
4. How many independent probability values do your network tables contain?

Bayesian Networks - Solution (2) I

1. *IcyWeather* is not caused by any of the car-related variables, so needs no parents. It directly affects the battery and the starter motor. *StarterMotor* is an additional precondition for Starts. The new network is shown in Figure:



Bayesian Networks - Solution (2) II

2. Reasonable probabilities may vary a lot depending on the kind of car and perhaps the personal experience of the assessor. The following values indicate the general order of magnitude and relative values that make sense:
   - A reasonable prior for *IcyWeather* might be 0.05 (perhaps depending on location and season).
   - $P(Battery|IcyWeather) = 0.95$, $P(Battery|\neg IcyWeather) = 0.997$.
   - $P(StarterMotor|IcyWeather) = 0.98$, $P(Battery|\neg IcyWeather) = 0.999$.
   - $P(Radio|Battery) = 0.9999$, $P(Radio|\neg Battery) = 0.05$.
   - $P(Ignition|Battery) = 0.998$, $P(Ignition|\neg Battery) = 0.01$.
   - $P(Gas) = 0.995$.
   - $P(Starts|Ignition, StarterMotor, Gas) = 0.9999$, other entries 0.0.
   - $P(Moves|Starts) = 0.998$.

Bayesian Networks - Solution (2) III

3. With 8 Boolean variables, the joint has $2^8 - 1 = 255$ independent entries
4. Given the topology shown in figure, the total number of independent CPT entries is $1 + 2 + 2 + 2 + 2 + 1 + 8 + 2 = 20$.

Bayesian Networks - Exercise (3) I

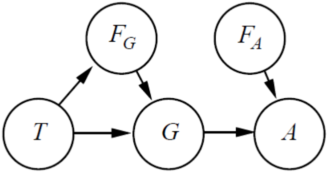
In your local nuclear power station, there is an alarm that senses when a temperature gauge exceeds a given threshold. The gauge measures the temperature of the core. Consider the Boolean variables A (alarm sounds), FA (alarm is faulty), and FG (gauge is faulty) and the multivalued nodes G (gauge reading) and T (actual core temperature).

1. Draw a Bayesian network for this domain, given that the gauge is more likely to fail when the core temperature gets too high.
2. Is your network a **polytree**? Why or why not?
3. Suppose there are just two possible actual and measured temperatures, normal and high; the probability that the gauge gives the correct temperature is x when it is working, but y when it is faulty. Give the conditional probability table associated with G.
4. Suppose the alarm works correctly unless it is faulty, in which case it never sounds. Give the conditional probability table associated with A.

5. Suppose the alarm and gauge are working and the alarm sounds. Calculate an expression for the probability that the temperature of the core is too high, in terms of the various conditional probabilities in the network.

Bayesian Networks - Solution (3) I

1. A suitable network is shown in figure below. The key aspects are: the failure nodes are parents of the sensor nodes, and the temperature node is a parent of both the gauge and the gauge failure node. It is exactly this kind of correlation that makes it difficult for humans to understand what is happening in complex systems with unreliable sensors.



2. No matter which way the student draws the network, it should not be a **polytree** because of the fact that the temperature influences the gauge in two ways.

3. The CPT for G is shown below

Table: CPT of G

	T=Normal		T=High	
	F_G	$\neg F_G$	F_G	$\neg F_G$
G=Normal	y	x	$1-y$	$1-x$
G=High	$1-y$	$1-x$	y	x

4. The CPT for A is as follows:

Table: CPT of A

	G=Normal		G=High	
	F_A	$\neg F_A$	F_A	$\neg F_A$
A	0	0	0	1
$\neg A$	1	1	1	0

Bayesian Networks - Solution (3) III

5. Abbreviating $T = High$ and $G = High$ by T and G, the probability of interest here is $P(T|A, \neg F_G, \neg F_A)$. Because the alarms behavior is deterministic, we can reason that if the alarm is working and sounds, G must be High. Because F_A and A are d-separated from T, we need only calculate $P(T|\neg F_G, G)$.
   The opportunistic way is to notice that the CPT entries give us $P(G|T, \neg F_G)$

   $$P(T|\neg F_G, G) \propto P(G|T, \neg F_G)P(T|\neg F_G)$$

   We then use Bayes Rule again on the last term:

   $$P(T|\neg F_G, G) \propto P(G|T, \neg F_G)P(\neg F_G|T)P(T)$$

   A similar relationship holds for $\neg T$:

   $$P(\neg T|\neg F_G, G) \propto P(G|\neg T, \neg F_G)P(\neg F_G|\neg T)P(\neg T)$$

Bayesian Networks - Solution (3) IV

Normalizing, we obtain

$$P(T|\neg F_G, G) =$$

$$\frac{P(G|T, \neg F_G)P(\neg F_G|T)P(T)}{P(G|T, \neg F_G)P(\neg F_G|T)P(T) + P(G|\neg T, \neg F_G)P(\neg F_G|\neg T)P(\neg T)}$$

Letting $P(T) = p$, $P(FG|T) = g$, and $P(FG|\neg T) = h$, we get

$$P(T|\neg F_G, G) = \frac{p(1-g)(1-x)}{p(1-g)(1-x) + (1-p)(1-h)x}$$

Decision Trees

Decision Tree Algorithms I

Consider the following data set comprised of three binary input attributes (A_1, A_2, and A_3) and one binary output:

Example	A_1	A_2	A_3	Output y
x_1	1	0	0	0
x_2	1	0	1	0
x_3	0	1	0	0
x_4	1	1	1	1
x_5	1	1	0	1

Table: samples with attributes

Use the following algorithm (copied from the textbook) to learn a decision tree for these data. Show the computations make to determine the attribute to split at each node.

Decision Tree Algorithms II

```
function DECISION-TREE-LEARNING(examples, attributes, parent_examples) returns a tree

  if examples is empty then return PLURALITY-VALUE(parent_examples)
  else if all examples have the same classification then return the classification
  else if attributes is empty then return PLURALITY-VALUE(examples)
  else
     A ← argmax_{a ∈ attributes} IMPORTANCE(a, examples)
     tree ← a new decision tree with root test A
     for each value v_k of A do
        exs ← {e : e ∈ examples and e.A = v_k}
        subtree ← DECISION-TREE-LEARNING(exs, attributes − A, examples)
        add a branch to tree with label (A = v_k) and subtree subtree
     return tree
```

Figure: Decision Tree Learning Algorithm. The function IMPORTANCE is described in Section 18.3.4 of textbook. The function PLURALITY-VALUE selects the most common output value among a set of examples, breaking ties randomly

Solution I

Note that to compute each split, we need to compute Remainder(A_i) for each attribute A_i, and select the attribute the provides the minimal remaining information, since the existing information prior to the split is the same for all attributes we may choose to split on.
Computations for first split: remainders for A_1, A_2, and A_3 are

Solution II

$$\left(\frac{4}{5}\right)\left(-\frac{2}{4}\times\log\left(\frac{2}{4}\right)-\frac{2}{4}\times\log\left(\frac{2}{4}\right)\right)+\left(\frac{1}{5}\right)\left(-0-\frac{1}{1}\times\log\left(\frac{1}{1}\right)\right)$$

$$=0.800$$

$$\left(\frac{3}{5}\right)\left(-\frac{2}{3}\times\log\left(\frac{2}{3}\right)-\frac{1}{3}\times\log\left(\frac{1}{3}\right)\right)+\left(\frac{2}{5}\right)\left(-0-\frac{2}{2}\times\log\left(\frac{2}{2}\right)\right)$$

$$\approx 0.551$$

$$\left(\frac{2}{5}\right)\left(-\frac{1}{2}\times\log\left(\frac{1}{2}\right)-\frac{1}{2}\times\log\left(\frac{1}{2}\right)\right)+$$

$$\left(\tfrac{3}{5}\right)\left(-\tfrac{1}{3}\times\log\left(\tfrac{1}{3}\right)-\tfrac{2}{3}\times\log\left(\tfrac{2}{3}\right)\right)\approx 0.951$$

Solution III

Choose A_2 for first split since it minimizes the remaining information needed to classify all examples. Note that all examples with $A_2 = 0$, are correctly classified as B = 0. So we only need to consider the three remaining examples (x_3, x_4, x_5) for which $A_2 = 1$.
After splitting on A_2, we compute the remaining information for the other two attributes on the three remaining examples (x_3, x_4, x_5) that have $A_2 = 1$. The remainders for A_1 and A_3 are

$$\left(\frac{2}{3}\right)\left(-\frac{2}{2} \times \log\left(\frac{2}{2}\right) - 0\right) + \left(\frac{1}{3}\right)\left(-0 - \frac{1}{1} \times \log\left(\frac{1}{1}\right)\right) = 0$$

$$\left(\frac{1}{3}\right)\left(-\frac{1}{1} \times \log\left(\frac{1}{1}\right) - 0\right) + \left(\frac{2}{3}\right)\left(-\frac{1}{2} \times \log\left(\frac{1}{2}\right) - \frac{1}{2} \times \log\left(\frac{1}{2}\right)\right) \approx 0.6$$

So, we select attribute A_1 to split on, which correctly classifies all remaining examples.

The End