Inference

09 January 2021 12:51

Bayesian probability
BBN and how to build them

Now we will discuss how to use inference in Bayesian Belief Networks

Probabilistic Inference in Belief Networks

Exact Inference

Probabilities are computed exactly

- a simple version of this is enumeration
- we can also use variable elimination, which is a method that exploits conditional independence

Approximate Inference

Approximates the probabilities and are characterized by different guarantees they provide

- they produce guaranteed bounds on the probabilities i.e. the exact probability will fall between a given range
- They may produce probabilistic bounds on the error i.e. the error is within 0.1 of the error 95% of the time. Such algorithms also guarantee that, as time increases, the probability estimates will converge to the exact answer.

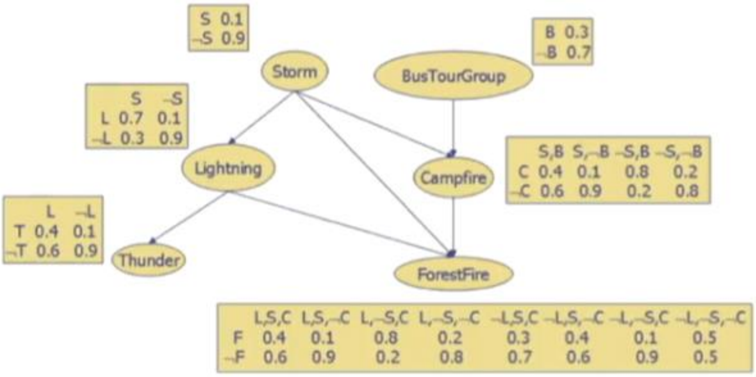
Type of reasoning

- Diagnostic reasoning - reasoning from symptom to cause
- Predictive Reasoning - reasoning from cause to symptom
- Intercausal Reasoning - reasoning about the mutual causes of a common effect
- Combined Reasoning - if the query variable is a parent of some observed variables and the descendent of other observed variables

Inference by Enumeration

- This involves enumerating through every world that is consistent with the evidence
- $p(X|e) = \alpha p(X|e) = \alpha \sum_y p(X, e, y)$
- $\alpha = 1/p(e) = 1/\sum_{x,y} p(X, e, y)$
- Remember that y is the set of hidden variables
    - These are variables in the network that are not included in the evidence or our query set, X.
    - If there are 2 hidden variables in the network, (A, B), we have to consider four different worlds.
    - $p(X|E \wedge A \wedge B)$
    - $p(X|E \wedge A \wedge \neg B)$
    - $p(X|E \wedge \neg A \wedge B)$
    - $p(X|E \wedge \neg A \wedge \neg B)$

Example



- Given the BBN pictured:
  - What is the probability of a forest fire, i.e. $X = F$, given that:
    - A Storm (S) took place
    - Lightning (L) was observed
    - No Bus Tour Group (B) visited the forest
- $p(F|S, L, \neg B) =?$

We have 3 values for our variables. We have 2 other variables that we don't know the value of.

If we sum the probabilities of all the possible resulting states, we can work out the probability of F.

The possible values of C and T are: C && T, C && ¬T, ¬C && T, ¬C && ¬T

- Calculate the posterior probability $p(F|S \wedge L \wedge \neg B)$
  - The event $E = S \wedge L \wedge \neg B \wedge F$
  - $\{S \wedge L \wedge \neg B \wedge C \wedge T \wedge F,$
    $S \wedge L \wedge \neg B \wedge C \wedge \neg T \wedge F,$
    $S \wedge L \wedge \neg B \wedge \neg C \wedge T \wedge F,$
    $S \wedge L \wedge \neg B \wedge \neg C \wedge \neg T \wedge F\}$
- $p(E) = \sum_{x \in E} p(x)$

So if we sum those states...

$$p(E) = \sum_{x \in E} p(x)$$

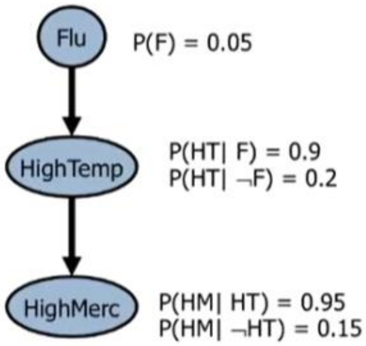
- $p(S \wedge \neg B \wedge L \wedge C \wedge T \wedge F)$
  $= p(S)p(\neg B)p(L|S)p(C|S \wedge \neg B)\ p(T|L)p(F|S \wedge L \wedge C)$
  $= 0.1 \times 0.7 \times 0.7 \times 0.1 \times 0.4 \times 0.4$
  $= 0.000784$
- $p(S \wedge \neg B \wedge L \wedge C \wedge \neg T \wedge F)$
  $= p(S)p(\neg B)p(L|S)p(C|S \wedge \neg B)\ p(\neg T|L)p(F|S \wedge L \wedge C)$
  $= 0.1 \times 0.7 \times 0.7 \times 0.1 \times 0.6 \times 0.4$
  $= 0.001176$
- $p(S \wedge \neg B \wedge L \wedge \neg C \wedge T \wedge F)$
  $= p(S)p(\neg B)p(L|S)p(\neg C|S \wedge \neg B)\ p(T|L)p(F|S \wedge L \wedge \neg C)$
  $= 0.1 \times 0.7 \times 0.7 \times 0.9 \times 0.4 \times 0.1$
  $= 0.001764$
- $p(S \wedge \neg B \wedge L \wedge \neg C \wedge \neg T \wedge F)$
  $= p(S)p(\neg B)p(L|S)p(\neg C|S \wedge \neg B)\ p(\neg T|L)p(F|S \wedge L \wedge \neg C)$
  $= 0.1 \times 0.7 \times 0.7 \times 0.9 \times 0.6 \times 0.1$
  $= 0.002646$

Note the evidence never changes, and F (our query variable) never changes.

Inference in a Chain of 3 Nodes

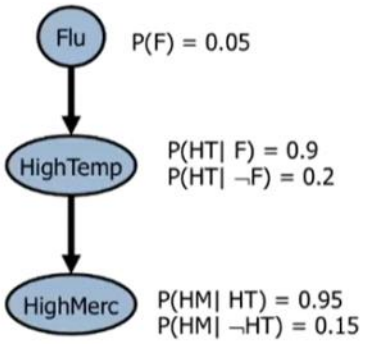
Example 1

- If Flu is Observed, we can use the chain rule: $p(HM|F) = \sum_{HighTemp} p(HM|HT)p(HT|F) = 0.9 \times 0.95 + 0.1 \times 0.15 = 0.87$
- Recalled that if A and C are conditionally independent: $p(C|A) = p(C|B)p(B|A) + p(C|\neg B)p(\neg B|A)$



Example 2

- If HighMecury readings are Observed, we can use Bayes rule and the Chain rule:
- $p(F|HM) = \alpha p(F)p(HM|F) = \alpha p(F) \sum_{HighTemp} p(HM|HT)p(HT|F) = \alpha \times 0.05 \times 0.87 = 0.0435\alpha$
- $p(\neg F|HM) = \alpha p(\neg F)p(HM|\neg F) = \alpha p(\neg F) \sum_{HighTemp} p(HM|HT)p(HT|\neg F) = \alpha \times 0.95 \times 0.31 = 0.2945\alpha$
- $\alpha = \frac{1}{0.0435+0.2945} \implies p(F|HM) = 0.1287$



Variable Elimination

Uses Bayesian Belief Networks. Another form of exact inference.

Adapted from a solution to finding solutions to CSP's.
Algorithm is based on the notion that a belief network specifies a factorisation of the joint probability distribution - this is more efficient than enumeration
Remember that a conditional probability is a function on a variable Y and some set of evidence variables
Also recall that the probability of the possible outcomes for Y, given a specific set of value assignments to the evidence, should sum to 1.
We may call a function on a set of variables a factor and we say that the scope of the factor is the set of variables it involves.
The conditional probability P(X|Y,Z) could be described as a factor with scope X, Y, Z

X	Y	$P(Z=t \mid X,Y)$
t	t	0.1
t	f	0.2
f	t	0.4
f	f	0.3

Expressing Factors

We can express out factors as an array
- if there is an ordering of the variables, e.g. alphabetical
- and the values in the domains are mapped into non-negative integers then
- there is a unique representation of each factor as a one-dimensional array

For example, we can represent the conditional probability table pictured as [0.1, 0.2, 0.4, 0.3]
We can perform a number of operations on factors - conditioning, summing and multiplying

Conditioning

If we have observed a variable, we can define a new factor with a new domain. The domain will be a subset of the domain prior to discovering the value of the variable.

As we discover more and more variables, the subset of possible values becomes smaller and smaller:

$r(X, Y, Z) =$

X	Y	Z	val
t	t	t	0.1
t	t	f	0.9
t	f	t	0.2
t	f	f	0.8
f	t	t	0.4
f	t	f	0.6
f	f	t	0.3
f	f	f	0.7

$r(X = t, Y, Z) =$

Y	Z	val
t	t	0.1
t	f	0.9
f	t	0.2
f	f	0.8

$r(X = t, Y, Z = f) =$

Y	val
t	0.9
f	0.8

$$r(X = t, Y = f, Z = f) = 0.8$$

Multiplying Factors

If two factors share a variable within their scope, we can multiply them together to make a new factors with a scope equivalent to the union of the original two factor's scope.

$f_1 =$

A	B	val
t	t	0.1
t	f	0.9
f	t	0.2
f	f	0.8

$f_2 =$

B	C	val
t	t	0.3
t	f	0.7
f	t	0.6
f	f	0.4

$f_1 * f_2 =$

A	B	C	val
t	t	t	0.03
t	t	f	0.07
t	f	t	0.54
t	f	f	0.36
f	t	t	0.06
f	t	f	0.14
f	f	t	0.48
f	f	f	0.32

Summing Factors

We eliminate a chosen variable, by adding together the outcomes for each possible value of the variable.
As with conditioning, it allows the removal of a variable in order to simplify our conditional table.

This is known as summing out a variable.

$f_3 =$

A	B	C	val
t	t	t	0.03
t	t	f	0.07
t	f	t	0.54
t	f	f	0.36
f	t	t	0.06
f	t	f	0.14
f	f	t	0.48
f	f	f	0.32

$\sum_B f_3 =$

A	C	val
t	t	0.57
t	f	0.43
f	t	0.54
f	f	0.46

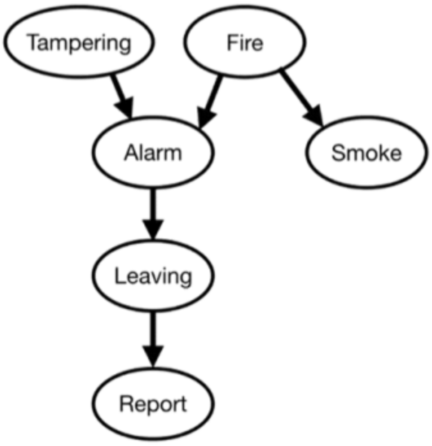
Back to Variable Elimination...

For a given query, we can represent the conditional probabilities of a BBN as a series of factors
The algorithm for solving a query is then as follow

- construct a factor for each conditional probability condition
- eliminate each non-query variable
    - if the variable is observed, its value is set to the observed value in each of the factors in which the variable appears
    - otherwise sum out the variable
- multiply the remaining factors and then normalise

Variable Elimination Example

- Consider the pictured network, it represents a number of conditional probabilities:
    - $p(Tampering), p(Fire), p(Alarm|Tampering, Fire)$ $p(Smoke|Fire), p(Leaving|Alarm), p(Report|Leaving)$
- How would we solve the query $P(Tampering|Smoke = True, Report = True)$?
- We perform step 1: construct our factors:
    - $f_0(Tampering), f_1(Fire), f_2(Alarm, Tampering, Fire)$ $f_3(Smoke, Fire), f_4(Leaving, Alarm),$ $f_5(Report, Leaving)$
- First, we set our observed variables, Smoke and Report, to their value, and eliminate them from the factors:
    - $f_0(Tampering), f_1(Fire), f_2(Alarm, Tampering, Fire)$ $f_3(Fire), f_4(Leaving, Alarm), f_5(Leaving)$



- Next, we eliminate the variable Fire.
- This involves multiplying together all factors that contain Fire, and then summing Fire out.
- $f_1(Fire) * f_2(Tampering, Alarm, Fire) * f_3(Fire) = f_{1,2,3}(Fire, Tampering, Alarm)$
- $\sum_{Fire} f_{1,2,3} = f_6(Tampering, Alarm)$

- We can repeat this same process with the factors containing Alarm, to eliminate that variable.
- We now have the following factors:
    - $f_0(Tampering), f_5(Leaving), f_7(Tampering, Leaving)$
- We're trying to find $Tampering$, so we must now eliminate $Leaving$
- Again, by multiplying and summing out $Leaving$, we have:
    - $f_0(Tampering), f_8(Tampering)$
- Finally, we multiply these factors together:
    - $f_9(Tampering) = f_0(Tampering) * f_8(Tampering)$
- Remember that although these factors have the same scope, their values will be different, as they are constructed using different evidence.

- Finally, we can find the posterior distribution over Tampering by:
- $f_9(Tampering) / \sum_{Tampering} f_9(Tampering)$
- The denominator here represents the prior probability of the evidence, in this case $Smoke = True, Report = True$.

Decision Making

- BBN's provide a mechanism to conduct Bayesian inference on large sets of random variables.
- These likelihood's of the outcomes must be used in some way to make a decision, a process not specifically support by a BBN. - for example, decide a treatment regime for a patient.
- We need to explicitly consider represent actions under consideration and utility of the resultant outcomes.
- We consider two possible tools for this decision making process
  - decision trees
  - influence diagrams, or decision networks

Expected Utility

- given a set of outcomes, of a particular action A, a utility function assigns a utility (value - measure of desirability) to each outcome
- Assuming that we have a probability distribution over the set of outcomes, the expected utility of taking the action A is defined as

- $EU(A) = \sum_i P(O_i|A) \times U(O_i|A)$

The assumption is that a Bayesian decision maker wants to maximise their expected utility through actions.

Combining utility Theory with Probability theory

- given evidence E, which action A is expected to deliver the most value
  - $EU(A|E) = \sum_i P(O_i|E, A) \times U(O_i|A)$

Utility can be a number of things, and is often represented as money

Outcomes

To inform our agent we must define relations between outcomes

- To inform our agent, we must define relations between outcomes.
- Consider two outcomes, σ_1 and σ_2
  - $\sigma_1 \succeq \sigma_2$ - σ_1 is **weakly preferred** to σ_2, meaning that σ_1 is at least as desirable as σ_2.
  - $\sigma_1 \sim \sigma_2$ - means that $\sigma_1 \succeq \sigma_2$ and $\sigma_2 \succeq \sigma_1$, and that the outcomes are equally preferred, meaning we are **indifferent** to which one we choose.
  - $\sigma_1 \succ \sigma_2$ - σ_1 is **strictly preferred** to σ_2 i.e. this means that we are not indifferent, and that we do not weakly prefer σ_2 to σ_1.
- These relations should be:
  - Complete - An agent has preferences between all pairs of outcomes
  - Transitive - If $\sigma_1 \succ \sigma_2$ and $\sigma_2 \succ \sigma_3$ then $\sigma_1 \succ \sigma_3$

Decision Trees

Decision Trees contain two types of nodes:

1. Chance nodes - represented by a circle
    a. representing a random variable
    b. edges emanating from change node represent the possible outcomes of the random variable and are labelled with the probability of the outcome
2. Decision node - represented by a square
    a. Representing a decision to be made
    b. edges emanating from such a node represent a set of mutually exclusive and exhaustive actions that the decision maker can make

The expected utility EU

- Of a change node is defined as the expected value of the utilities associated with its outcomes
- Of a decision alternative (action) is the expected utility of the chance node encountered if the action is taken
- Of a decision node is the maximum utility of all its alternatives

Normative theory - the assumption that the agent wants to maximise expected utility
Prospect theory - the final end state is not what people have preferences over - rather what matters is how much the choice differs from the current situation - 10k means more to someone with $500 than someone with $100,000.

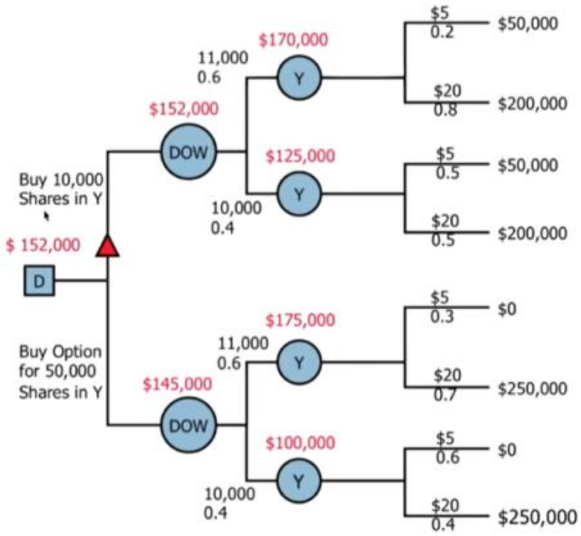
Representing Risk Aversion

Include it in the utility function

- One way to model an individual's attitude to risk is with a utility function that maps money to utility
    - For example using a exponential utility function, where R is the risk tolerance. The larger the value of R, the more risk-tolerant the user
    - $U_R(x) = 1 - e^{-x/R}$
    - For R = 500:
        - $EU(BuyX) = 0.25 * U_{500}(500) + 0.25 * U_{500}(1000) + 0.5 * U_{500}(2000) = 0.865$
        - $EU(Bank) = U_{500}(1050) = 0.877$
    - For R = 1000:
        - $EU(BuyX) = 0.25 * U_{1000}(500) + 0.25 * U_{1000}(1000) + 0.5 * U_{1000}(2000) = 0.688$
        - $EU(Bank) = U_{500}(1050) = 0.65$

A more complex example

- Sue is considering buying 10,000 shares in company Y @ $10/share
    - Her buying this number of shares will affect the price of Y
- She believes that the overall value of the DOW industrial average will also affect the price of Y
    - She believes that, in one month the DOW will either be at 10,000 or 11,000
    - And that Y will be at $5 or $20 per share
- Alternatively, she can buy an option of Y worth $100,000
    - Allows her to buy 50,000 shares in Y for $15/share in one month
- Her beliefs are:
    - $p(DOW = 11,000) = 0.6$
    - $p(Y = \$5 | Decision = buy, DOW = 11,000) = 0.2$
    - $p(Y = \$5 | Decision = buy, DOW = 10,000) = 0.5$
    - $p(Y = \$5 | Decision = option, DOW = 11,000) = 0.3$
    - $p(Y = \$5 | Decision = option, DOW = 10,000) = 0.6$



Influence Diagrams

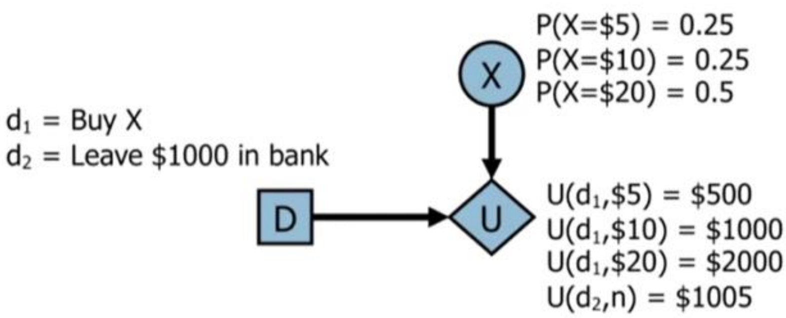
- Decision trees are fine but grow exponentially
- Influence diagrams suffer from neither of these issues
- Contains 3 nodes: chance (circle), decision (square), and utility (diamond)

A decision network (aka infuence diagram) is a graphical representation of a finite sequential decision problem. They extend belief networks to include decision variables and utility.
A decision network extends the single-stage decision network to allow for sequential decisions and allows both chance nodes to be parents of decision nodes.

Different edges have different meanings
- Edge to chance node - value of node is probabilistically dependent on the value of the parent
- Edge to decision node: value of the parent is known at the time the decision is made
  - If parent is a decision node, the edge represents a decision sequence
- Edge to a utility node - value of node is deterministically dependent on the value of the parent

The chance nodes satisfy the **Markov Condition**

These diagrams are *Bayesian Networks augmented with a utility node and ordered decision nodes.*



- Solving an influence diagram
  - Which decision choice has the maximum utility? i.e. $max(EU(d_1), EU(d_2))$
  - $EU(d_1) = E(U|d_1) = P(X = \$5) \times U(d_1, \$5) + P(X = \$10) \times U(d_1, \$10) + P(X = \$20) \times U(d_1, \$20) = \$1375$
  - $EU(d_2) = E(U|d_2) = \$1005$
- The decision is d_1.

Evaluating Influence Diagrams

with 1 d node:

1. add any evidence (set probability of value of random variables observed to 1)
2. for each action value in the decision node
   a. set decision node to that value
   b. calculate posterior probabilities of nodes that are parent of the utility node
   c. calculate the expected utility of the action
3. Return action with highest utility

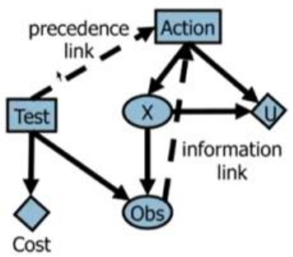
Information Links

Links from chance nodes to decision nodes
Indicate that the chance node must be known before a decision is made corresponding to that decision node
= can be used to explicitly calculate what decision should be made, given the different values for a chance node

- Evaluating Influence Diagram with Information Links
  - Add any evidence (set probability of value of random variables observed to 1)
  - For each combination of values of the parents of the decision node, for each action value in the decision node:
    - Set decision node to that value
    - Calculate posterior probabilities of nodes that are parents of the utility node
    - Calculate the expected utility of the action
  - Record the resulting expected utility for the action
  - Return the table of actions and associated expected utility values (decision table)

Sequential Decision Making

- Typical Example is a Test-Action Influence Diagram
  - Test Decision Node must be evaluated first
  - The cost associated with a test is included as a separate utility node
  - If the decision is to run a test, evidence will be obtained as a result of the test
    - The chance node representing the observation, Obs, has an information link to the Action decision
    - One value, unknown, of Obs will represent the decision to not test within CPT
    - $p(Obs = unknown | Test = no) = 1$
    - $p(Obs = unknown | Test = yes) = 0$



- Evaluating such an influence diagram:
  - Evaluate decision network with any available evidence
  - Enter test decision as evidence
  - If test decision is not 'yes', use value 'unknown'
  - Evaluate Action decision

Summary

- In this topic, we have covered:
  - Definitions of probability functions, spaces, distributions and propositions
  - Probabilistic inference
  - Theorems of probability, including Bayes Rule
  - Bayesian Belief Networks, their construction and use
  - Variable elimination
  - Utility functions
  - Influence networks, their definition and how to evaluate them