

DTAD-20: Advanced Artificial Intelligence: Bayesian Reasoner Homework

Purpose: Gain hands-on experience with Bayesian Networks.

1 Assignment

1. Exercise 14.1(a-d) on page 533 of the textbook.
2. Exercise 14.3(c,d) on page 534
3. Exercise 14.4 on page 534.
4. Implement the Enumerate-Ask and Enumerate-All algorithms on page 506 of the textbook.

2 Further Details on Enumerate-Ask

Inputs:

1. A data set
2. The variables of a boolean network and their legal values.
3. The topology (i.e., parent-child connections) for the boolean network.

Your system should read in a data set (D), preferably from a file. It should also read in the variables being used and their possible values. Finally, the topology of the Bayesian Network should be read in: i.e. which nodes are the parents of which other nodes.

Given the input network topology, your system should know WHICH conditional probability tables (CPTs) that need to be created. The system should then go through the data set and use it as the basis for computing the necessary entries in the CPTs (as we discussed in the lectures).

Your system will then accept queries about the probability distributions of a particular variable, given one or more assigned evidence variables, such as:

$$P(\textit{World - Economy} \mid \textit{Bush - wins} = \textit{True} \wedge \textit{Global - Warming} = \textit{True}) \quad (1)$$

In other words, find the conditional probabilities of the different values for the future of the world economy (i.e., Better, Worse, Same), given that George Bush wins reelection in 2004 and that global warming is actually happening. Here, World-Economy is the query variable, and Bush-wins and Global-Warming are the evidence variables.

The Enumerate-All subroutine will use the CPTs to answer these queries. It should be able to answer any query involving the domain variables, as long as only 1 query variable is used at a time. Extending to multiple query variables is not that difficult, so you might try that if you have extra time and/or interest.

Verify that your system works using TWO DOMAINS of your choosing. Feel free to make up one of the domains and data sets, but use a real data set (see the UC Irvine Machine-Learning Repository) for the other domain. This is a course in AI algorithms, not political science, medicine, auto repair, etc, so any data set is fine as long as it has at least 5 attributes and 20 or more atomic events (i.e. instances). Be prepared to explain the code and the domains to your thick-skulled instructor during the demo session.