CMPE 540 Assignment #2
"A Rule Based Bayesian Medical Expert"

Problem Statement

Assume that you want to build a system that is an expert in the field of medical diagnosis. If you get a medical encyclopedia and look up the entry, say, influenza, you will find that all of the symptoms are given, and that there is not any argument about these symptoms. In other words, given the symptoms, an accurate diagnosis could be made every time.

Using information organized in this way however, suggests that what you should do is to pick up a patient, decide he has influenza, then look in the encyclopedia to see if he has the right symptoms. However you should actually pick up a patient, decide what the symptoms are, and then look up to see what he is suffering from.

You will develop an expert system that will use a Bayesian inferencing system. We provide you a Knowledge Base for the domain of Medical Diagnosis with knowledge on nearly 100 different disease types and their diagnoses. The figures given are, roughly accurate.

Organization of Data Files

The data are given in two files:
Illnesses file **illness.txt**, and symptoms file **symptoms.txt.** The structure of these files is explained below.

Illnesses file

The structure of **illness.txt** is as follows:

```
Ilnessname p,(j,py,pn,)999
```

It contains an outcome and all the information relating to it. The first is the name of the outcome- say INFLUENZA. The next item, **p**, is the prior probability of that outcome *P(H)* -- this is the probability of this outcome occurring given no further information at all. We then have a series of repeated items with three elements. The first element, *j*, is the number of relevant symptom (or variable if you want to call it something else). The next two items are *P(E|H)* and *P(E|not H)* - the probability of getting a **Yes** answer to this variable given the outcome is true and the probabilities of getting a **Yes** answer if it is not true. The last item, 999 indicates the end of one particular illness. For example:

```
INFLUENZA, 0.01,1,0.9,0.01,2,1,0.01,3,0,0.01,999
```

This says that, in the case of influenza, there is prior probability *P(H)=0.01* of any random person having this illness. Now suppose that the program asks question 1 (symptom 1). We have *P(E|H)=0.9* and *P(E|not H)=0.01* which says that if the patient has influenza then nine times out of ten he will answer **Yes** to this question and, if he does not have influenza, then he will only answer **Yes** in one case in a hundred. Obviously, a **Yes** answer supports the hypothesis that he has influenza. A **No** answer tends to suggest that he has not.

Now consider, the second symptom/probability group *(2,1,0.01)*. In this case *P(E|H)=1* which says that, if he has influenza, then he must have this symptom. He might have the symptom without influenza *(P(E|not H)=0.01)* but it is not very likely.

Question 3 rules out influenza if he gives a **Yes** answer because *P(E|H)=0*. This could have been a question like: *Have you had the symptoms for most of your life*?

Symptoms file

The structure of symptoms.txt is as follows:

```
NUMBER, QUESTION
```

The first number is the number of the symptom, and QUESTION is the question to be asked to the patient to determine whether he has the symptom or not.

An example is

```
1, ARE YOU SNEEZING A LOT?
```

What to do

Develop the expert system with the above knowledgebase using CLIPS expert system shell (available at http://www.ghg.net/clips/CLIPS.html). Use Bayesian decision theory for uncertainty management. The expert system based on the answers of the user will try to diagnose the illness.

What to Submit

a) A hardcopy report that discusses your knowledge representation and transcripts of several consultations (some successful, some unsuccessful).

b) Attached to the report a CD containing

   a. your source codes,

b. sample input and corresponding output files and

Deadline

November 27, 2006 Monday 17:00