CPS331 Lecture: Coping with Uncertainty; Discussion of Dreyfus Reading

January 28, 2014

Objectives:

1. To discuss ways of handling uncertainty (probability; Mycin CF)
2. To discuss Dreyfus' views on "expert" systems

Materials:

1. Slide show of Green / Blue Cab Problem
2. Dreyfus Discussion Question

I. **Coping With Uncertainty**

A. We saw earlier that knowledge representation is a key problem in AI. One issue in knowledge representation poses a particular challenge for work in expert systems, and hence was discussed in Cawsey's discussion of Mycin: the problem of <u>uncertain</u> knowledge.

1. In the case of factual knowledge, we may think a certain fact is true, but may not be totally sure of it. For example, a medical diagnostic system may need to know the date on which a certain symptom first appeared, but the patient may only be able to say "I think it was last Friday, but I'm not sure."

2. Further, a given heuristic rule in an expert system may work most of the time, but not always. For example, a given set of symptoms may indicate a certain disease 90% of the time, but not always. (We will see later that this was a key issue in the development of the MYCIN system.)

a) In particular, uncertain rules often arise in expert systems that do diagnosis because of the need to do <u>abductive</u> reasoning.

(1) It is often the case that, given a problem, we may list symptoms that are always present when the problem is present. That is, we can say:

problem $\rightarrow$ symptoms

and if we know the problem, we can infer the symptoms with a high degree of certainty, using modus ponens.

(2) However, diagnostic systems are required to reason the other way - from symptoms to the problem. This is abductive reasoning, which is not sound. The best that can be said is that when certain symptoms are present, a given problem is a likely cause.

Example: Suppose a person walks into the room with wet hair. What are you likely to think?

ASK

Certainly, if it is raining heavily outside, someone walking to class without an umbrella will get we hair. But wet hair might also be caused by a person rushing out of the shower to get to class, or someone pulling a prank with a squirt gun, or ...

b) However, this is not the only source of uncertainty. To go back to our example of problems and symptoms, in medicine it is often the case that certain symptoms of a given disease appear in most, but not all, of the patients with that disease. Thus, even the inference

problem $\rightarrow$ symptom

is uncertain.

B. Several different approaches have been explored to reasoning when facts and/or rules involve uncertainty. We will look at two.

1. An approach based on probabilities and Bayes' theorem.

2. The approach used by MYCIN

C. One approach to uncertainty is based on a branch of mathematics known as probability theory. This has the advantage of bringing years of work in mathematics to bear on the problem. Unfortunately. it also has a major practical limitation.

1. Instead of representing facts by boolean predicates, we may represent them by probabilities - with a 0 meaning that a given statement is known to certainly be false, a 1 meaning that it is known to certainly be true, and a value between 0 and 1 expressing different degrees of likelihood that it is true.

   a) We denote the a-priori probability of some statement S by P(S). The a-priori probability of a statement is simply the fraction of the cases in which the statement will be true in a random sample

      Example: The a-priori probability that a randomly chosen person is male is about 1/2, so we may write P(male) = 0.5

   b) We also must deal with conditional probabilities - i.e. if we already have some evidence about a situation, then the probability of a certain outcome may differ.

      Example: The conditional probability that a randomly chosen Gordon student is male is only about 1/3, so we write P(male | Gordon student) = 0.33

2. Similarly, we may associate a probability with a rule, expressing the probability that the conclusion is true, given the truth of the premises.

   Example: Consider the following automobile diagnosis rule

   if   the car won't start
             the starter does not turn
             the lights don't come on

   then   replace the battery

Suppose that 2% of the time, when all three premises are true, the problem turns out to be something other than the battery. Then we would associate a probability of 98% with this rule.

i.e. P(battery | ¬start ∧ ¬starter turns ∧ ¬lights) = 0.98

3. In the case of expert systems doing diagnostic reasoning, it turns out that the probabilities we need are often not directly available.

   Example: Suppose we wanted to associate a probability with the rule

   if the patient has a fever
      the patient has red spots
   then the patient has measles

   a) What we need to know is "in a population of individuals having a fever and red spots, what fraction have measles?" This data may not be readily available, however.

   b) There is a way to calculate this, though, using a theorem known as Bayes' theorem and more readily available information

      (1) Bayes' theorem says

      $$P(H \mid E) = \frac{P(E \mid H) * P(H)}{P(E)}$$

      i.e. the probability of some hypothesis (e.g. the patient has measles) given some evidence (e.g. the patient has a fever and red spots) is equal
      • to the probability of the evidence given the hypothesis (e.g. that a patient with measles will have a fever and red spots)
      • times the intrinsic probability of the hypothesis (e.g. that a person chosen at random has measles)
      • divided by the intrinsic probability of the evidence (e.g. that a person chosen at random has a fever and red spots)

(2) To use Bayes' theorem in this case, we would need to know

ASK

(a) The fraction of patients with measles who have a fever

(b) The fraction of patients with measles who have red spots

(c) The fraction of people in general who have measles

(d) The fraction of people in general who have a fever

(e) The fraction of people in general who have red spots

(f) The fraction of people who have a fever who also have red spots

(This example illustrates that a large amount of information may be necessary to apply probability theory rigorously.)

(3) An example of using Bayes' theorem

(a) Consider the following story (from the portion of teh Dreyfus chapter you were not required to read, but may have chosen to read)

"Two cab companies operate in a city, the Blue and the Green, the names coming from the colors of their cabs. Of all taxis, 85 percent are Blue, and 15 percent are Green. One night, a taxi was involved in a hit-and-run accident. No information is available about how many cabs of each color were on duty that night or what cabs were in that part of the city at the time of the accident. A witness identified the cab as Green. The court tested the

witness's ability to distinguish between Blue and Green cabs at night under the same conditions as were present at the time of the accident and found that the witness was able to identify the color of a cab correctly 80 percent of the time and was wrong 20 percent of the time. What are the chances that that the cab involved in the accident was Green as claimed by the witness?

(b) How many of you have had a chance to try this mental experiment as Dreyfus suggested? What did you conclude?

ASK

PROJECT and work through slide show

(c) Recall that Bayes' theorem tells us

$$P(H \mid E) = \frac{P(E \mid H) * P(H)}{P(E)}$$

where, in this case H (the hypothesis) is that the cab was actually green, and E (the evidence) is that the witness would say it was green

i) $P(E \mid H)$ is simply the probability that the witness will say the cab is Green if it is. We are told that the witness's statements are true 80% of the time, so we take this to be 0.8.

ii) $P(H)$ is simply the a-priori probability that the cab is Green. Given that only 15% of the cabs are Green, (and assuming that there is no difference in accident rate between Green and Blue cabs), we take this to be 0.15.

iii) $P(E)$ is the a-priori probability that the witness says that a given cab is Green. We are not told this, but can calculate it from:

P(says Green)

= P(says Green ^ is Green) + P(says Green ^ is not Green)

= P (says Green | is Green) * P (is Green) + P (says Green | ¬ Green) * P (¬ Green)

= 0.8 * 0.15 + 0.2 * 0.85 = 0.12 + 0.17 = 0.29

    (d) Hence P (H | E)   = 0.8 * 0.15 / 0.29 = .41 (as Dreyfus claimed in the chapter!)

D. Instead of using true probabilities, expert systems may use <u>confidence levels.</u>

   1. These look a bit like probabilities, but they are not.

   2. We will consider here the approach taken by MYCIN, which has been followed by many other systems. This approach is sometimes called the Stanford Certainty Factor Algebra, because MYCIN was developed at Stanford.

     a) MYCIN used an ad-hoc approach based on confidence levels, with no attempt to treat them rigorously as probabilities.

     b) Further, instead of ranging from 0 .. 1, like probabilities, MYCIN's confidence levels ranged from -1 to +1, with -1 meaning "I am certain this is not the case" and +1 meaning "I am certain this is the case". A confidence level of 0 would mean "I have no idea whether this is true or false."

     Note: MYCIN actually records two measures of confidence for any piece of information - a measure of belief (MB) and a measure of disbelief (MD). [Generally one of these is 0.] The confidence factor is defined as MB - MD.

     c) In forming conclusions, MYCIN treated any conclusion with certainty -0.2 .. +0.2 as unsupported in either the positive or negative direction; -1 .. -0.2 as probably false, and +0.2 .. +1 as

probably true. If a rule requires that a certain antecedent hold, then it will only fire if the certainty associated with that antecedent is at least 0.2 (i.e. -1.0 .. 0.2 is regarded as not true.) Likewise, if a rule requires that a certain antecedent not hold, then it will fire only if the certainty associated with that antecedent is <= -0.2 (i.e -0.2 .. 1.0 is regarded as not false.)

d) For output to the user, the certainty factor was converted to an appropriate English qualifier such as "there is weak evidence that ..." or "there is suggestive evidence that ..." or "there is strong evidence that ..."

3. In MYCIN, both facts and rules can have certainty factors associated with them. In particular, rules normally qualify their conclusion by a certainty factor, like the following actual MYCIN rule.

if: the stain of the organism is gramneg
   the morphology of the organism is rod
   the patient is a compromised host

then: conclude that the identity of the organism is pseudomonas
   with confidence 0.6.

4. MYCIN combined certainties as follows:

a) The certainty of the conclusion of a rule is the certainty associated with the rule itself (e.g. 0.6 in the case cited above) * the MINIMUM certainty of any antecedent. (But note that if any antecedent has certainty < 0.2, then the rule won't fire at all.)

b) However, the way the antecedents are phrased affected this. Some comparison functions that can be used in stating antecedents return a numeric value between -1 and 1, while others simply report either false or true (the former if the numeric certainty is < 0.2, the latter if it is $>= 0.2$). If an antecedent is phrased using one of these functions, and it tests

true, it will count as 1.0 in determining the combined certainty for the rule.

   c) If a rule reaches a conclusion that has already been produced by some other rule (i.e. the conclusion is an OR node), then the new certainty is

   C1 + C2 * (1 - C1) -- where C1 is the old certainty
   and C2 is the certainty factor from the new rule.

   Thus, if several separate lines of reasoning lead to the same conclusion, then the confidence in that conclusion is strengthened - but can never exceed 1.0.

E. The approaches to reasoning with uncertain information we have considered are but two of many. This remains an open research question, where researchers seek to find a system that strikes a good balance between:

   1. Theoretical rigor

   2. Practical usefulness

II. Discussion of Dreyfus Reading (do all but last as entire class if time is short - last certainly do in groups)

A. For the two tasks at the start of the chapter

   1. For which task(s) did you give the correct answer?

   2. For which task(s) did you give the incorrect answer?

   3. Which task did you find harder? Why?

B. The Dreyfus brothers draw a distinction between "knowing that" and "knowing how". What is the distinction? What examples do they cite to illustrate the difference between the two?

(Answers from book)

1. Driving a shift car - suddenly thinking about rules for the correct gear.

2. AF flight evaluator suddenly confronted with an emergency in a plane he is an expert on after getting out of practice.

3. A/D 4/7 game vs check approval rule (formally same, but one much harder)

C. Dreyfus titled this chapter "Five Steps from Novice to Expert".   What are the five steps?   What characterizes each?   What role do rules and "deliberative rationality" play at each step?

(Answers from book)

1. Novice (context free rules)

2. Advanced Beginner (situational rules)

3. Competent (sense of goals; what is important; "problem-solving" approach)

4. Proficient (intuition instead of just deliberation)

5. Expertise (fluidity; arationality)

D. Does what Dreyfus describes ring true in your own experience of developing expertise in various areas?

E. Dreyfus describes the relationship between rationality and expertise by using the term "arationality". What does he mean by this?

F. Discussion question for small groups: What would Dreyfus say is the highest level of expertise a computer program based on rule-following could attain? Do you agree or disagree? Why?