GKMCET

Lecture Plan

Subject code & Subject Name: CS2351 & AI **Unit Number: 1V**

UNIT-IV: UNCERTAINTY

Uncertainty

[To act rationally under uncertainty we must be able to evaluate how likely certain things are.

[With FOL a fact F is only useful if it is known to be true or false.

[But we need to be able to evaluate how likely it is that F is true.

[By weighing likelihoods of events (probabilities) we can develop mechanisms for acting rationally under uncertainty.

Dental Diagnosis example.

[In FOL we might formulate

[P. symptom(P,toothache)→

disease(p,cavity) disease(p,gumDisease)

disease(p,foodStuck) L

[When do we stop?

[Cannot list all possible causes.

[We also want to rank the possibilities. We don't want to start drilling for a cavity before checking for more likely causes first.

Axioms Of Probability

[Given a set U (universe), a probability function is a function defined over the subsets of U that maps each subset to the real numbers and that satisfies the Axioms of Probability

1.Pr(U) = 1

2.Pr(A) [0,1]

3.Pr(A B) = Pr(A) + Pr(B) –Pr(A ∩B)

Note if A ∩B = {} then Pr(A B) = Pr(A) + Pr(B)

BASIC PROBABILTY NOTATION

1 Unconditional or prior probabilities

2 Conditional or posterior probabilities

SEMANTICS OF BAYESIAN NETWORK

1 Representation of joint probability distribution

2 Conditional independence relation in Bayesian network

INFERENCE IN BAYESIAN NETWORK

1 Tell

2 Ask

3 Kinds of inferences

4 Use of Bayesian network

TEMPORAL MODEL

1 Monitoring or filtering

2 Prediction

Bayes' Theorem

Many of the methods used for dealing with uncertainty in expert systems are based on Bayes' Theorem.

Notation:

P(A) Probability of event A

P(A B) Probability of events A and B occurring together

P(A | B) Conditional probability of event A given that event B has occurred

If A and B are independent, then P(A | B) = P(A).

Expert systems usually deal with events that are not independent, e.g. a disease and its symptoms are not independent.

Theorem

P (A B) = P(A | B)* P(B) = P(B | A) * P(A) therefore P(A | B) = P(B | A) * P(A) / P(B

Uses of Bayes' Theorem

In doing an expert task, such as medical diagnosis, the goal is to determine identifications (diseases) given observations (symptoms). Bayes' Theorem provides such a relationship.

P(A | B) = P(B | A) * P(A) / P(B)

Suppose: A = Patient has measles, B = has a rash
Then:P(measles/rash)= P(rash/measles) * P(measles) / P(rash)

The desired diagnostic relationship on the left can be calculated based on the known statistical quantities on the right.

Joint Probability Distribution

Given a set of random variables X_1 ... X_n, an atomic event is an assignment of a particular value to each X_i.

The joint probability distribution is a table that assigns a probability to each atomic event. Any question of conditional probability can be answered from the joint.[Example from Russell & Norvig.]

	Toothache	¬ Toothache
Cavity	0.04	0.06
¬ Cavity	0.01	0.89

Problems:

- The size of the table is combinatoric: the product of the number of possibilities for each random variable.
- The time to answer a question from the table will also be combinatoric.
- Lack of evidence: we may not have statistics for some table entries, even though those entries are not impossible.

Chain Rule

We can compute probabilities using a chain rule as follows:

P(A &and B &and C) = P(A | B &and C) * P(B | C) * P(C)

If some conditions C_1 &and ... &and C_n are independent of other conditions U, we will have:

P(A | C_1 &and ... &and C_n &and U) = P(A | C_1 &and ... &and C_n)

This allows a conditional probability to be computed more easily from smaller tables using the chain rule.

Bayesian Networks

Bayesian networks, also called belief networks or Bayesian belief networks, express relationships among variables by directed acyclic graphs with probability tables stored at the nodes.[Example from Russell & Norvig.]

1 A burglary can set the alarm off
2 An earthquake can set the alarm off
3 The alarm can cause Mary to call
4 The alarm can cause John to call

Computing with Bayesian Networks

If a Bayesian network is well structured as a poly-tree (at most one path between any two nodes), then probabilities can be computed relatively efficiently.

One kind of algorithm, due to Judea Pearl, uses a message-passing style in which nodes of the network compute probabilities and send them to nodes they are connected to.

Several software packages exist for computing with belief networks.

A Hidden Markov Model (HMM) tagger chooses the tag for each word that maximizes: [Jurafsky, op. cit.] P(word | tag) * P(tag | previous n tags)

For a bigram tagger, this is approximated as: $t_i = \text{argmax}_j P(w_i | t_j) P(t_j | t_{i-1})$

In practice, trigram taggers are most often used, and a search is made for the best set of tags for the whole sentence; accuracy is about 96%.

GKMCET

Lecture Plan

Subject code & Subject Name: CS2351 & AI **Unit Number: 1V**