Chapter (13)

This chapter gives initial intuitions on how an agent can handle uncertainty. Uncertainty happens when an agent may never know for certain what state it's in or where it will end up after a sequence of actions. This may happen because of different reasons such as:

1. **Partial Observability**: interpreting partial sensor information, a logical agent must consider every logically possible explanation for the observations, no matter how unlikely. This leads to impossible large and complex belief-state representations.
2. **Nondeterminism**: Environment may act differently given some observed evidence as they are not enough to correctly identify the agent's state.
3. **Laziness:** It is too much work to list the complete set of antecedents or consequents needed to ensure an exceptionless rule.
4. **Theoretical ignorance**: No complete theory for the domain.

Probabilities as a degree of belief representation

Any agent, based on the reasons stated above, can have a degree of belief in its state instead of 100% belief. Probability provides the best way of summarizing such uncertainty because

1. Probabilities are numeric values, so they describe exactly how the world behaves based on the agent's knowledge base—rather than being just descriptions of an observer's degree of belief. Probabilities are transitive and comparable.
2. Probabilities are updatable, so they are a flexible representation of the agent's degree of belief as evidence arrives online (i.e. Bayesian View).
3. Probability calculations can be represented in a vectorized format and all probability rules will still hold, so it allows a faster computational way of representing multiple propositions.

Decision Theory

Probability is not enough to take decisions. The agent needs to know which states can get the best outcome (Utility Theory), so that it can take actions that may get to the best state with a high probability (get the Maximum Expected Utility 'MEU'), so decision theory is a combination of both utility and probability theories.

Some Definitions of Probability Theory

We introduce some definitions and rules of thumb in probability theory. A possible world is an assignment of values to all of the random variables. Sample Space is the set of all possible worlds. Possible worlds are mutually exclusive and exhaustive. Prior probability refers to degrees of belief in propositions in the absence of any evidence. Evidence is some known information that can limit out the sample space. Posterior probability refers to degrees of belief given some evidence. When making decisions, an agent needs to condition on all the evidence it has observed which may be infeasible if there are a lot of pieces of evidence (scalability problem). A probability model is completely determined by a full joint probability distribution (the joint distribution for all of the random variables) because every proposition's probability is a sum over possible worlds. Full joint probability distribution grows exponentially with the number of random variables (scalability problem).

Conditional Probability as a matter of interest

We are interested in computing conditional probabilities of some variables, given evidence about others. Conditional probabilities can be found by first using joint probability

$$P(a,\ b)\ =\ P(a|b)\ .\ P(b) \tag{1-a}$$

$$1/P(b) \text{ is just a normaliztion factor, let it be } \alpha$$
$$P(a|b) = \alpha \, . \, P(a, b) \tag{1-b}$$

then using general marginalization rule for any sets of variables Y and Z

$$P(Y) = \sum_{z \in Z} P(Y, z) \tag{2}$$

If the query involves a single variable, 'X', let 'E' be the list of evidence variables, 'e' be the list of observed values for them, and 'Y' be the remaining unobserved variables. from equations (1-c) and (2)

$$P(X|e) = \alpha \, . \, P(X, e) = \alpha \, . \sum_{y \in Y} P(X, e, y)$$

where P(X,e,y) can be evaluated directly from the full joint distribution as the variables X, E, and Y constitute the complete set of variables.

<u>Scalability Problem and Independence</u>

The problem in this approach is all about scalability, as the complexity of computing 'P(X,e,y)' on all combinations of possible values for each 'y' in 'Y' grows exponentially with the number of variables in 'Y'. It's not logical to compute all of these probabilities as a lot of variables in our experiment may be completely independent (i.e. no covariance), or conditionally independent that, based on domain knowledge, they don't affect each other, but some cause may affect both of them, so they appear statistically dependent. Absolute and conditional independence can change the complexity from exponential to linear, as the full joint distribution can be factored into separate joint distributions.

$$\text{conditional independence}: P(Effect1, Effect2, \ldots \mid Cause) = \prod_i P(Effecti|Cause) \tag{1}$$

$$\text{absolute independence}: \; P(Cause|Effect1, \, Effect2) = P(Cause|Effect1) \tag{2}$$
$$\text{in the case that } Effect2 \text{ cannot affect the Cause}$$

<u>Bayes' Rule</u>

Bayes' rule is a simple equation underlies most probabilistic inference systems derived from (1)

$$P(Y|X, \, e) = \frac{P(X|Y, \, e) \, . \, P(Y|e)}{P(X|e)} \tag{3}$$

in case of simple cause-effect relationships, Bayes' Rule will look like that

$$P(cause|effect) = \frac{P(effect|cause) \, . \, P(cause)}{P(effect)} \tag{4}$$

where P(cause|effect) is called diagnostic, and P(effect|cause) is called causal.

Bayes' Rule is important because

1. It helps to evaluate diagnosis based on causal and prior knowledge of cause existence, as diagnostic knowledge is often more fragile than causal knowledge, as causal knowledge is less likely to change online than diagnostic knowledge.
2. Bayes' rule combines evidence as much as possible to make posterior evaluation much simpler using absolute and conditional independence in equations (4) and (5).