Tecnológico de Monterrey

Hidden Markov Models

Daniel Otero Fadul

Science Department
School of Engineering and Science

Probabilistic Reasoning over Time

Let us assume that we see the world in a discrete way: we “take photos” of reality every certain amount of time T_m, also known as *sampling time*, and we observe what is recorded in these photos; some things can be observed while others remain hidden.

Let $t_k = kT_m$, $k = 0, 1, 2, \ldots$. Let X_k be the *state of the system* at time t_k, which is a set of variables that are not observable, and let E_k be the evidence at this same time t_k, which is the set of variables that we can observe.

In what follows, we will use the following notation: $X_{m:n} = X_m, X_{m+1}, \ldots, X_n$, $m, n \in \{0, 1, 2, \ldots\}$, $m < n$.

Transition Model

When we include time in our models, it is customary to say that the probability distribution of the current state X_k given the information of the previous states $X_{0:k-1}$ is equal to

$$P(X_k|X_{0:k-1}).$$

This is known as the *transition model*.

Markov Processes

Note that in the above expression, it is stated that the current state X_k depends on all previous states $X_{0:k-1}$. A very common assumption is that the current state X_k depends only on a finite number of its previous states. This is called the *Markov assumption*. All systems that satisfy this property are called **Markov processes**. For example, the simplest version of the Markov assumption would be the following:

$$P(X_k|X_{0:k-1}) = P(X_k|X_{k-1}).$$

This variant is known as a *first-order Markov process*.

Another important assumption is that the form of the probability distribution does not change for each time t_k, but is always the same. A stochastic process that satisfies this property is called a **stationary process**.

Sensor Model

Regarding the evidence E_k, which is related to the *sensor model*, we have the following *Markov assumption for the sensor*:

$$P(E_k | X_{0:k}, E_{1:k-1}) = P(E_k | X_k).$$

The term $P(E_k|X_k)$ is known as the *sensor model*. Note that it is assumed that there is no evidence for time t_0.

Hidden Markov Models

On the other hand, the probability distribution of the initial state is denoted as $P(X_0)$. Taking the above into account, the joint probability distribution of states $X_{0:k}$ and evidence $E_{1:k}$ is equal to

$$P(X_{0:k}, E_{1:k}) = P(X_0) \prod_{i=1}^{k} P(X_i|X_{i-1}) P(E_i|X_i). \tag{1}$$

The three terms on the right-hand side of the expression are known as the initial state model $P(X_0)$, the transition model $P(X_i|X_{i-1})$, and the sensor model $P(E_i|X_i)$.

An Example

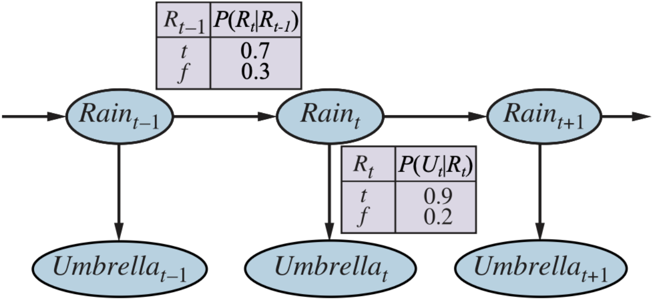

Figura: Structure of the Bayesian network and conditional probability distributions in the umbrella example. The transition model is $P(\text{Rain}_k|\text{Rain}_{k-1})$ and the sensor model is $P(\text{Umbrella}_k|\text{Rain}_k)$. Image taken from [1].

Inferences

Given a generic temporal model, there are four basic types of inferences that we can perform:

- **Filtering:** This type of inference is known as *state estimation*. In mathematical terms, this is equivalent to obtaining the probability distribution of $P(X_k|E_{1:k} = e_{1:k})$.
- **Prediction:** The objective of this inference is to calculate the posterior probability distribution of a future state given all the evidence collected: $P(X_{k+j}|E_{1:k} = e_{1:k})$, $j > 0$.
- **Smoothing:** In this case, we want to calculate the probability distribution of a past state X_j given all the evidence: $P(X_j|E_{1:k} = e_{1:k})$ para algún j tal que $0 \leq j \leq k$.
- **Most likely explanation:** Given a sequence of observations, we want to obtain the most probable sequence of states that generated these observations:

$$\max_{x_{1:k}} P(x_{1:k}|E_{1:k} = e_{1:k}).$$

Filtering

Let us consider the previous example for a moment. Suppose we want to calculate $P(R_2|U_{1:2} = [t,t])$. Then, using equation (1), we have that

$$P(R_2|U_{1:2} = [t,t]) = \alpha \sum_{R_1} \sum_{R_0} P(U_2 = t|R_2)P(R_2|R_1)P(U_1 = t|R_1)P(R_1|R_0)P(R_0),$$

where α is a normalization constant.

BIBLIOGRAFÍA

1 Stuart Rusell,Peter Norvig, S. J., *"Artificial Intelligence: A Modern Approach"*, Fourth Edition, Prentice Hall, 2020.