CMSC 170 - Introduction to Artificial Intelligence
Revised 2017

Hidden Markov Models

A **Hidden Markov Model** is a statistical model where a system is represented as a simple Bayesian model called a **Markov Chain**. In a Markov Chain, illustrated in Fig. 1., the **next state is solely dependent on the current state**; it is independent of any past and future states.

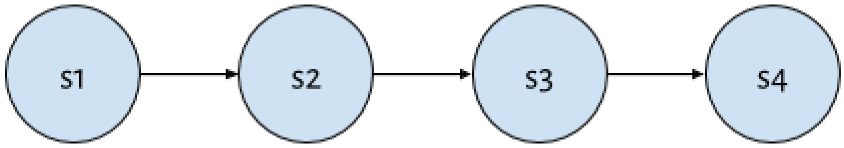

Fig. 1. A Markov Chain.

Thus, from Fig. 1., s2 is dependent on s1, s3 is dependent on s2, but not s1, and so on. In a Hidden Markov Model, the Markov Chain is **hidden**, that is, its state cannot be directly observed. Instead, each state has an **observable measurement value**, as shown in Fig. 2.

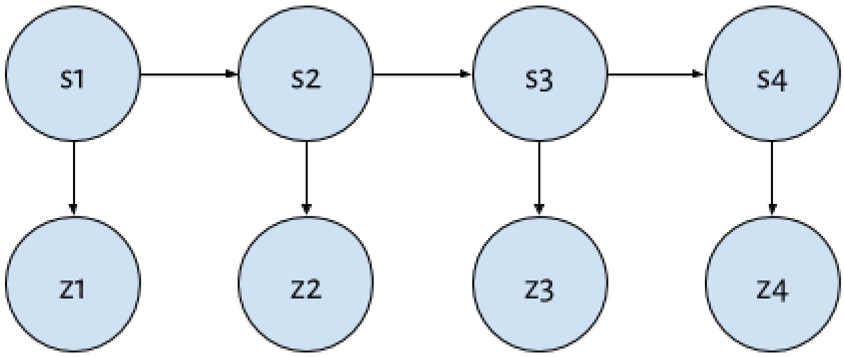

Fig. 2. A Markov Chain (above) with each state's measurement values (below).

A Markov Chain can be described by a Bayes network as follows: given that the value of a state can be either S or T, Fig. 3 shows the possible states and transitions from each state.

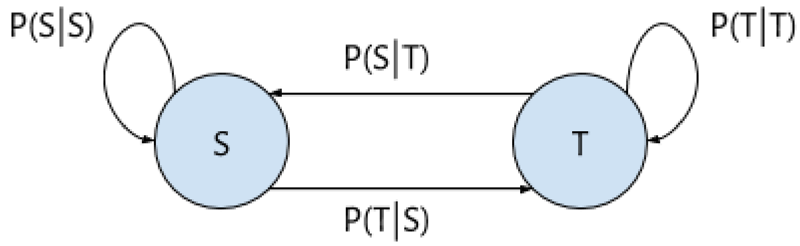

Fig. 3. The Bayes network for a Markov Chain whose state can either be S or T.

Solving for the Transition Probabilities

The transition probabilities can be solved if a sequence of states from the Markov Chain are given. For example, given the sequence

STSSTSTSSSTT

The following probabilities can be computed:

Probability	Value	Explanation
P(S0)	1	The sequence starts with an S.
P(S|S)	3/7	Out of 7 occurrences of S with a next state, 3 were followed by an S.
P(T|S)	4/7	Out of 7 occurrences of S with a next state, 4 were followed by a T.
P(T|T)	1/4	Out of 4 occurrences of T with a next state, 1 was followed by a T.
P(S|T)	3/4	Out of 4 occurrences of T with a next state, 3 were followed by an S.

Predicting the Probability Value of the Next State

Once the probability value of the first state is computed (e.g., P(S0)), it can be used to predict the probability values of future states using total probability:

$$P(S_n) = P(S_n|S_{n-1})P(S_{n-1}) + P(S_n|T_{n-1})P(T_{n-1})$$

or

$$P(T_n) = P(T_n|S_{n-1})P(S_{n-1}) + P(T_n|T_{n-1})P(T_{n-1})$$

For example, given the probability values computed in the previous section, we can compute P(T1) as:

$$P(T_1) = P(T_1|S_0)P(S_0) + P(T_1|T_0)P(T_0)$$
$$P(T_1) = (4/7) \times 1 + 0.25 \times 0$$
$$P(T_1) = 0.57142857142$$

Further computations of future states yield:

$$P(T_2) = P(T_2|S_1)P(S_1) + P(T_2|T_1)P(T_1)$$
$$P(T_2) = 0.57142857142 \times 0.42857142857 + 0.25 \times 0.57142857142$$
$$P(T_2) = 0.24489795918 + 0.14285714285 = 0.38775510203$$

Using Measurement Values

Now, given observable measurement values that have a dependence on each state of the Markov Chain (as illustrated in Fig. 2.), we can solve for the probability value of the current state in the Markov Chain. For example, given the transition diagram for the Markov Chain, and the relationship between each possible state and the possible measurement values, E and F, both illustrated in Fig. 4.

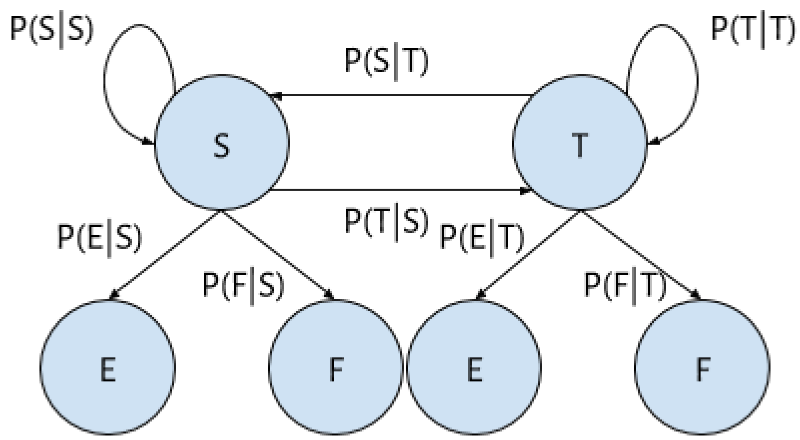

Fig. 4. Transition diagram for possible values a state, S and T, and the possible measurement values per state.

We can compute for the P(S1|E1) using Bayes Rule:

$$P(S_1|E_1) = \frac{P(E_1|S_1)P(S_1)}{P(E_1)}$$

P(E1|S1) is given in the transition diagram, and P(S1) can be computed as discussed earlier. P(E1) can be expanded using total probability as:

$$P(E_1) = P(E_1|S_1)P(S_1) + P(E|T_1)P(T_1)$$

Further states with their corresponding probability values can be computed in the same way. For example, to compute P(S3|E3):

$$P(S_3|E_3) = \frac{P(E_3|S_3)P(S_3)}{P(E_3)}$$

The only new computation here is P(S3), which is computed by using P(S2), which in turn, is computed using P(S1). We have already showed how P(S1) is computed earlier.

Exercise

Given an input file, **hmm.in**, with the following format:

```
S T //the possible values for each state in the Markov Chain
E F // the possible observable measurement values for each state in the Markov Chain
P(E|S) P(F|S) //a pair of values for P(E|S) and P(F|S), respectively
P(E|T) P(F|T) //a pair of values for P(E|T) and P(F|T), respectively
STSSTSTSSSTT //sequence used to derive transition probability values
3 //number of cases
S1 given E1 //compute P(S1|E1)
T3 given F3 //compute P(T3|F3)
S2 given F2 //compute P(S2 given F2)
```

There will be no UI option for this exercise. Write all results to another file, hmm.out with the following format:

```
P(S1 given E1) = //insert answer here
P(T3 given F3) = //insert answer here
P(S2 given F2) = //insert answer here
```

The scoring will be as follows:

Criteria	Score
Read input file	2
Compute transition probability values	5
Compute P(Sn) or P(Tn) values	2
Compute P(En) or P(Fn) values	2
Compute P(Sn|En) (and other similar values)	1
Write output file	1
Total	13

Bonus: Apply Laplace smoothing, and **add a field for *k* at the beginning of hmm.in**(2pts).