Introduction to Hidden Markov Model

Alexandre Savard
Schulich School of Music - McGill University
555 Sherbrooke St. West
Montreal, QC Canada H3A 1E3

***Abstract*— This paper introduces the fundamental concepts surrounding Hidden Markov Model. Readers who have never been introduced to this theory will find particularly useful the information provided by this short introduction before getting involved in more extensive explanations. It is in fact a summary of a paper written by Rabiner (1989) that is strongly recommended by the author. Markov process will first be explained before being extendd to the class of hidden Markov model. Finally, the three fundamental problems that can be solved by HMM will be presented.**

***Index Terms*— Hidden Markov Model, HMM**

I. Introduction

PEOPLE concerned in physics and engineering are probably more used to think about signal in terms of continuous function as audio or electromagnetic waves. However, the definition of signal appears to be larger, including discrete codes as alphabet or traditional music notation. Any types of signal can be represented using a model.

Models are very powerful tools used mostly for prediction or identification task but also for recognition systems. These models can be divided into two different classes: the deterministic models and the statistical models. The first class are the ones that we are the more use to deal with in sciences. Two well-known applications of this modeling approach are the one dimensional wave equation as well as the harmonic pendulum equation. The other type of model involves stochastic process including Gaussian process, Poisson process, Markov process and hidden Markov model. What make this last one very specific is that it handle sequences of consecutive observations while most of the other types of model don't.

II. Discrete Markov Process

Consider a set of N distinct states $\{S_1, S_2, ...S_N\}$. The system undergoes a change of state according to a set of probabilities. Each change corresponds to a time instant $t = 1, 2, ...$ The actual state at time t is denoted q_t. Such a model is described by figure 1. The key element of the Markov process is the definition of the probabilities of changes for an new state in function of previous states. A first order Markov chain is a probabilistic description involving the actual state and the previous one.

$$a_{ij} = P[q_t = S_j | q_{t-1} = S_i] \quad 1 \leq i, j \leq N$$

a_{ij} are the state transition probabilities with condition:

$$a_{ij} \geq 0$$

$$\sum_{j=1}^{N} a_{ij} = 1.$$

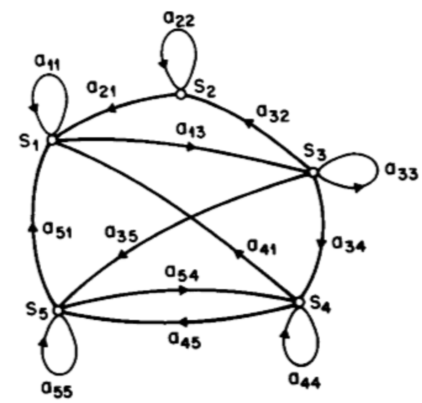


Fig. 1. A Markov chain with 5 states (labeled S_1 to S_5) with selected state transitions.

As an example, let consider the weather. We assume three different states corresponding to rainy, cloudy and sunny meteorological conditions. The transition probabilities matrix describes how the weather changes.

$$A = \{a_{ij}\} = \begin{pmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{pmatrix}$$

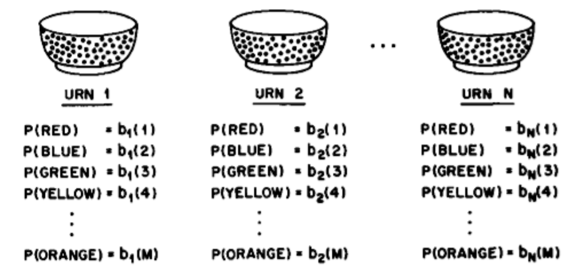
The values on the primary diagonal describe the probabilities to get the exact same weather two successive days. If you provide an observation sequence $O = \{S_3, S_3, S_3, S_1, S_1, S_3, S_2, S_3, \}$, you can calculate the probability to get this specific set of observation according to the model as following:

$$\begin{aligned} P(O|Model) &= P[S_3, S_3, S_3, S_1, S_1, S_3, S_2, S_3] \\ &= P[S_3] \bullet P[S_3|S_3] \bullet P[S_3|S_3] \bullet ... \bullet P[S_3, S_2] \\ &= \pi_3 \bullet a_{33} \bullet a_{33} \bullet a_{31} \bullet a_{11} \bullet a_{11} \bullet a_{13} \bullet a_{32} \bullet a_{23} \\ &= 1 \bullet (0.8)(0.8)(0.1)(0.4)(0.3)(0.1)(0.2) \\ &= 1.536 \times 10^{-4} \end{aligned}$$

III. Extension to Hidden Markov Model

So far we considered Markov models that only that only take into account states that were easily associated to observable physical events. An extension to hidden Markov model includes the cases where the observations are function of the state. We then obtain two different levels of stochastic process, the first one for the states, the other of the observations at each states.

The urn ball model shown in figure 2 depicts this situation. We pick up balls of different colors from different urns. Each urn corresponds to a specific state. Different observation probabilities coresponding to the color of the balls are associated to states.



However, and here is the key idea behind hidden Markov model, the observation doesn't provide us the state. The states are actually the hidden part of the model.

$$O = \{\text{green}, \text{red}, \text{blue}, ...\}$$

To be completely defined, the number of states N of the model has to be given as well as the number of observation symbols M per state. Also, transition state probability distribution $A = \{a_{i,j}\}$ and observation symbol probability distribution $B = \{b_j\}$ have to be specified. Finally, to complete the model, we need the initial state distribution $\pi = \{\pi_i\}$where $\pi_i = P[q_1 = S_i]$. The model is then represented by $\lambda = \{A, B, \pi\}$.

IV. HMM fundamental problems

There is mostly three problems that can be solved using hidden Markov models :

1) Given the observation sequence O and a model λ, how do we efficiently compute $P(O|\lambda)$, the probability of the observation sequence according to the model ?
2) Given the observation sequence O and a model λ, how do we choose a corresponding state sequence Q which is optimal in some meaningful sense (best explains the observations) ?
3) How do we adjust the model parameters λ to maximize $P(O|\lambda)$?

V. Conclusion

Hopefully, this paper will make easier the understanding of more extended and advanced papers related to hidden Markov model. This model is very important in the field machine learning algorithm and is commonly encountered. Its extension to musical application is recent. However, its use for audio analysis is appropriate since it has been first designed for this task in the field of speech recognition.

References

[1] Rabiner, L. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77. 257–85