

# 1 Hidden Markov Models

Hidden Markov models are one of the most widely used and best known variants of probabilistic finite automata [11, 10]. The theoretical basis for hidden Markov models and associated methods and algorithms were first described in a series of works by L. E. Baum et al. [5, 4, 7, 6, 3].

One can view a hidden Markov model as an extension of a standard Markov chain. A Markov chain based model has the property that every observed symbol corresponds directly to an associated state of the model. This property is too restrictive for numerous problems. Hidden Markov models therefore introduce the concept of a hidden (unobservable) states with a probability distribution over the observable symbols. The outputted symbol thus becomes a probabilistic function of the current hidden state. This extension allows application of hidden Markov models to a much larger variety of problems, however also poses new complications, such as increased complexity of evaluation of probability of a given signal (sequence of observable symbols), determining the optimal sequence of hidden states for a given signal or estimation of parameters for the model [10].

To better illustrate the mechanics behind the hidden Markov model we provide a simple example. Consider a lottery game with several urns filled with balls of different colours. Every urn may hold different number of balls of a certain colour, or may not even contain balls of some colours at all. In every turn an unbiased arbiter selects one urn randomly (but possibly abiding to some rules given by the urn selected previously) and takes out a random ball out of the selected urn. You are then shown the colour of the ball, not however, from which urn it was taken. The ball is then returned to the appropriate urn and the game continues so forth your goal being to predict the colours that come next. In this simple case, you can observe a sequence of symbols - colours generated by what can be viewed as a hidden Markov model, where the urns are the hidden states and the observable symbol probability distribution for every urn (state) is given by the colours of the balls inside.

More formally, a finite discrete hidden Markov model with  $n \in \mathbb{N}$  (hidden) states  $S = \{S_1, S_2, \dots, S_n\}$  over an alphabet of  $m \in \mathbb{N}$  observable symbols  $\Sigma = \{\sigma_1, \sigma_2, \dots, \sigma_m\}$  is a tuple:

$$\lambda = (A, B, \pi)$$

Where:

$A$  is an  $n$  times  $n$  square matrix such that an element of the matrix  $a_{ij} \in [0, 1]$  represents the transition probability from state  $S_i$  to state  $S_j$ . Naturally this implies  $\forall i \in \{1, 2, \dots, n\} : \sum_{j=1}^n a_{ij} = 1$ .

$B$  is an  $n$  times  $m$  matrix such that an element of the matrix  $b_{ij} \in [0, 1]$  represents the probability of outputting symbol  $\sigma_j$  in state  $S_i$ . Naturally this implies  $\forall i \in \{1, 2, \dots, n\} : \sum_{j=1}^m b_{ij} = 1$ .

$\pi$  is a vector of  $n$  variables,  $\pi = (\pi_1, \dots, \pi_n) \in [0, 1]^n$ . Where the value  $\pi_i$  represents the probability of state  $S_i$  being the initial state. Naturally this implies  $\sum_{i=1}^n \pi_i = 1$

Such a hidden Markov model can be used to generate a sequence of observable symbols (signal):

$$O = (o_1, \dots, o_T)$$

Where  $\forall t \in \{1, \dots, T\} : o_t \in \Sigma$ ,  $o_t$  is the symbol observed at time  $t$  and  $T$  is the number of discrete time steps during the observation (number of observed symbols).

Similarly, we denote a sequence of hidden states of the hidden Markov model as:

$$Q = (q_0, \dots, q_T)$$

Where  $\forall t \in \{1, \dots, T\} : q_t \in S$ ,  $q_t$  is the state of the model at time  $t$  and  $T$  is the number of visited states.

Any signal  $O$  generated by the hidden Markov model has a corresponding sequence of hidden states the model visited during the generation  $Q$  of the same length. This sequence is typically unknown for modeled real world signals and generally numerous different hidden state sequences may generate the observed signal with different probabilities.

It should be noted that the hidden Markov model defined here is finite in the sense of both  $n$  and  $m$  being finite numbers. An extension to model allowing infinite number of states, symbols or both is rather straightforward, however we will not be needing this extension for the purposes of this publication and thus it will be omitted.

## 1.1 Signal Observation Probability

## 1.2 Estimating Model Parameters

## 1.3 Classification

In general case, the hidden Markov model as defined would be ergodic - between each couple of states, there exists a finite path with non-zero probability. For many applications it is desirable to restrict the model in some fashion. The most commonly known of such restrictions is the so called

left-right model also known as the Bakis model. This model only allows transitions in the hidden state space in one direction, formally:

$$\forall i, j \in \{1, \dots, n\}; i < j : a_{ji} = 0$$

The left-right model is used extensively for speech recognition. [2, 9].

The hidden Markov model considered here follows the standard definition where the observable symbol is outputted in a state of the model. A modification is possible and used, again in the field of speech recognition, in which the outputted observable symbol is associated with a transition instead of a state [10, 1]. It has been proven useful to incorporate transitions that output no symbol (null transitions) into this modification [1].

The hidden Markov model defined in this section is a discrete hidden Markov model, with both discrete probability distribution of the observable symbols as well as a discrete distribution of transition probabilities on hidden states. It is possible to generalise a hidden Markov model by changing both of the afore mentioned probability distributions into continuous ones. A hidden Markov model with both continuous distribution on observable symbols and on hidden states is commonly known as a Bayesian network [8].

As was mentioned before, the definition we presented is for a finite model but can be extended to account for infinite number of states or observable symbols straightforwardly. Many other variants, modifications and extensions also exist, some of them hinted in [10].

## References

- [1] L. R. Bahl, F. Jelinek, and R. Mercer, “A maximum likelihood approach to continuous speech recognition,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PAMI-5, no. 2, pp. 179–190, March 1983.
- [2] R. Bakis, “Continuous speech recognition via centisecond acoustic states,” *Acoustical Society of America Journal*, vol. 59, p. 97, Jan. 1976.
- [3] L. Baum, “An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes,” *Inequalities*, vol. 3, pp. 1–8, 1972.
- [4] L. E. Baum and J. A. Eagon, “An inequality with applications to statistical estimation for probabilistic functions of markov processes and to a model for ecology,” *Bulletin of the American Mathematical*

- Society*, vol. 73, no. 3, pp. 360–363, 05 1967. [Online]. Available: <http://projecteuclid.org/euclid.bams/1183528841>
- [5] L. E. Baum and T. Petrie, “Statistical inference for probabilistic functions of finite state markov chains,” *The Annals of Mathematical Statistics*, vol. 37, no. 6, pp. 1554–1563, 12 1966. [Online]. Available: <http://dx.doi.org/10.1214/aoms/1177699147>
  - [6] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, “A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains,” *The Annals of Mathematical Statistics*, vol. 41, no. 1, pp. 164–171, 02 1970. [Online]. Available: <http://dx.doi.org/10.1214/aoms/1177697196>
  - [7] L. E. Baum and G. R. Sell, “Growth transformations for functions on manifolds,” *Pacific Journal of Mathematics*, vol. 27, no. 2, pp. 211–227, 1968. [Online]. Available: <http://projecteuclid.org/euclid.pjm/1102983899>
  - [8] I. Ben-Gal, *Bayesian Networks*. John Wiley & Sons, Ltd, 2008. [Online]. Available: <http://dx.doi.org/10.1002/9780470061572.eqr089>
  - [9] F. Jelinek, “Continuous speech recognition by statistical methods,” *Proceedings of the IEEE*, vol. 64, no. 4, pp. 532–556, April 1976.
  - [10] L. R. Rabiner, “A tutorial on hidden markov models and selected applications in speech recognition,” in *PROCEEDINGS OF THE IEEE*, 1989, pp. 257–286.
  - [11] S. Verwer, R. Eyraud, and C. de la Higuera, “Pautomac: a pfa/hmm learning competition,” Tech. Rep., February 2011.