INTERNATIONAL COMPUTER SCIENCE INSTITUTE

1947 Center St. • Suite 600 • Berkeley, California 94704-1198 • (510) 643-9153 • FAX (510) 643-7684

Markov Models and Hidden Markov Models: A Brief Tutorial

Eric Fosler-Lussier

TR-98-041

December 1998

Abstract

This tutorial gives a gentle introduction to Markov models and Hidden Markov models as mathematical abstractions, and relates them to their use in automatic speech recognition. This material was developed for the Fall 1995 semester of *CS188: Introduction to Artificial Intelligence* at the University of California, Berkeley. It is targeted for introductory AI courses; basic knowledge of probability theory (*e.g.* Bayes' Rule) is assumed. This version is slightly updated from the original, including a few minor error corrections, a short "Further Reading" section, and exercises that were given as a homework in the Fall 1995 class.

1 Markov Models

Let's talk about the weather. Here in Berkeley, we have three types of weather: *sunny*, *rainy*, and *foggy*. Let's assume for the moment that the weather lasts all day, i.e. it doesn't change from rainy to sunny in the middle of the day.

Weather prediction is all about trying to guess what the weather will be like tomorrow based on a history of observations of weather. Let's assume a simplified model of weather prediction: we'll collect statistics on what the weather was like today based on what the weather was like yesterday, the day before, and so forth. We want to collect the following probabilities:

$$P(w_n \mid w_{n-1}, w_{n-2}, \ldots, w_1) \tag{1}$$

Using expression 1, we can give probabilities of types of weather for tomorrow and the next day using n days of history. For example, if we knew that the weather for the past three days was {sunny, sunny, foggy} in chronological order, the probability that tomorrow would be rainy is given by:

$$P(w_4 = Rainy \mid w_3 = Foggy, w_2 = Sunny, w_1 = Sunny) \tag{2}$$

Here's the problem: the larger n is, the more statistics we must collect. Suppose that $n = 5$, then we must collect statistics for $3^5 = 243$ past histories. Therefore, we will make a simplifying assumption, called the *Markov Assumption*:

In a sequence $\{w_1, w_2, \ldots, w_n\}$:

$$P(w_n \mid w_{n-1}, w_{n-2}, \ldots, w_1) \approx P(w_n \mid w_{n-1}) \tag{3}$$

This is called a *first-order* Markov assumption, since we say that the probability of an observation at time n only depends on the observation at time $n-1$. A *second-order* Markov assumption would have the observation at time n depend on $n-1$ and $n-2$. In general, when people talk about Markov assumptions, they usually mean first-order Markov assumptions; I will use the two terms interchangeably.

We can the express the joint probability using the Markov assumption.[1]

$$P(w_1, \ldots, w_n) = \Pi_{i=1}^{n} P(w_i \mid w_{i-1}) \tag{4}$$

So this now has a profound affect on the number of histories that we have to find statistics for— we now only need $3^2 = 9$ numbers to characterize the probabilities of all of the sequences. This assumption may or may not be a valid assumption depending on the situation (in the case of weather, it's probably not valid), but we use these to simplify the situation.

So let's arbitrarily pick some numbers for $P(w_{\text{tomorrow}} \mid w_{\text{today}})$, expressed in Table 1.

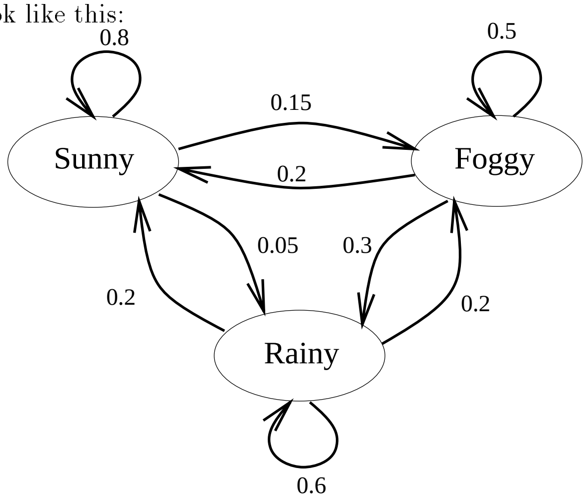
		Tomorrow's Weather		
		Sunny	Rainy	Foggy
Today's Weather	Sunny	0.8	0.05	0.15
	Rainy	0.2	0.6	0.2
	Foggy	0.2	0.3	0.5

Table 1: Probabilities of Tomorrow's weather based on Today's Weather

For first-order Markov models, we can use these probabilities to draw a probabilistic finite state automaton. For the weather domain, you would have three states (Sunny, Rainy, and Foggy), and

[1] One question that comes to mind is "What is w_0?" In general, one can think of w_0 as the $START$ word, so $P(w_1|w_0)$ is the probability that w_1 can start a sentence. We can also just multiply the prior probability of w_1 with the product of $\Pi_{i=2}^{n} P(w_i \mid w_{i-1})$; it's just a matter of definitions.

every day you would transition to a (possibly) new state based on the probabilities in Table 1. Such an automaton would look like this:



1.1 Questions:

1. Given that today is sunny, what's the probability that tomorrow is sunny and the day after is rainy?

   Well, this translates into:

$$
\begin{aligned}
P(w_2 = \text{Sunny}, w_3 = \text{Rainy}|w_1 = \text{Sunny}) &= P(w_3 = \text{Rainy} \mid w_2 = \text{Sunny}, w_1 = \text{Sunny}) * \\
&\quad P(w_2 = \text{Sunny} \mid w_1 = \text{Sunny}) \\
&= P(w_3 = \text{Rainy} \mid w_2 = \text{Sunny}) * \\
&\quad P(w_2 = \text{Sunny} \mid w_1 = \text{Sunny}) \\
&= (0.05)(0.8) \\
&= 0.04
\end{aligned}
$$

   You can also think about this as moving through the automaton, multiplying the probabilities as you go.

2. Given that today is foggy, what's the probability that it will be rainy two days from now?

   There are three ways to get from foggy today to rainy two days from now: {foggy, foggy, rainy}, {foggy, rainy, rainy}, and {foggy, sunny, rainy}. Therefore we have to sum over these paths:

$$
\begin{aligned}
P(w_3 = \text{Rainy} \mid w_1 = \text{Foggy}) &= P(w_2 = \text{Foggy}, w_3 = \text{Rainy} \mid w_1 = \text{Foggy}) + \\
&\quad P(w_2 = \text{Rainy}, w_3 = \text{Rainy} \mid w_1 = \text{Foggy}) + \\
&\quad P(w_2 = \text{Sunny}, w_3 = \text{Rainy} \mid w_1 = \text{Foggy}) + \\
&= P(w_3 = \text{Rainy} \mid w_2 = \text{Foggy}) P(w_2 = \text{Foggy} \mid w_1 = \text{Foggy}) + \\
&\quad P(w_3 = \text{Rainy} \mid w_2 = \text{Rainy}) P(w_2 = \text{Rainy} \mid w_1 = \text{Foggy}) + \\
&\quad P(w_3 = \text{Rainy} \mid w_2 = \text{Sunny}) P(w_2 = \text{Sunny} \mid w_1 = \text{Foggy})
\end{aligned}
$$

$$
\begin{aligned}
&= (0.3)(0.5) + (0.6)(0.3) + (0.05)(0.2) \\
&= 0.34
\end{aligned}
$$

Note that you have to know where you start from. Usually Markov models start with a null start state, and have transitions to other states with certain probabilities. In the above problems, you can just add a start state with a single arc with probability 1 to the initial state (sunny in problem 1, foggy in problem 2).

2 Hidden Markov Models

So what makes a Hidden Markov Model? Well, suppose you were locked in a room for several days, and you were asked about the weather outside. The only piece of evidence you have is whether the person who comes into the room carrying your daily meal is carrying an umbrella or not.

Let's suppose the following probabilities:

	Probability of Umbrella
Sunny	0.1
Rainy	0.8
Foggy	0.3

Table 2: Probabilities of Seeing an Umbrella Based on the Weather

Remember, the equation for the weather Markov process before you were locked in the room was:

$$P(w_1, \ldots, w_n) = \Pi_{i=1}^{n} P(w_i \mid w_{i-1}) \tag{5}$$

Now we have to factor in the fact that the actual weather is *hidden* from you. We do that by using Bayes' Rule:

$$P(w_1, \ldots, w_n \mid u_1, \ldots, u_n) = \frac{P(u_1, \ldots, u_n | w_1, \ldots, w_n) P(w_1, \ldots, w_n)}{P(u_1, \ldots, u_n)} \tag{6}$$

where u_i is true if your caretaker brought an umbrella on day i, and false if the caretaker didn't. The probability $P(w_1, \ldots, w_n)$ is the same as the Markov model from the last section, and the probability $P(u_1, \ldots, u_n)$ is the prior probability of seeing a particular sequence of umbrella events (e.g. {True, False, True}). The probability $P(u_1, \ldots, u_n | w_1, \ldots, w_n)$ can be estimated as $\Pi_{i=1}^{n} P(u_i | w_i)$, if you assume that, for all i, given w_i, u_i is independent of all u_j and w_j, for all $j \neq i$.

2.1 Questions:

1. Suppose the day you were locked in it was sunny. The next day, the caretaker carried an umbrella into the room. Assuming that the prior probability of the caretaker carrying an umbrella on any day is 0.5, what's the probability that the second day was rainy?

$$
\begin{aligned}
P(w_2 = \text{Rainy}| \\ w_1 = \text{Sunny}, u_2 = \text{True}) &= \frac{P(w_2 = \text{Rainy}, w_1 = \text{Sunny} | u_2 = \text{T})}{P(w_1 = \text{Sunny} | u_2 = \text{T})} \\
(u_2 \textit{ and } w_1 \textit{ independent}) &= \frac{P(w_2 = \text{Rainy}, w_1 = \text{Sunny} | u_2 = \text{T})}{P(w_1 = \text{Sunny})}
\end{aligned}
$$

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | **sales@docketalarm.com** | **1-866-77-FASTCASE**