Markov Models and Hidden Markov Models: A Brief Tutorial

Yibin Jiang
Definition:

a stochastic model used to model randomly changing systems where it is assumed that future states depend only on the present state and not on the sequence of events that preceded it.
Weather Prediction

Question:

To guess what the weather will be like tomorrow based on a history of observations of weather.
Weather Prediction

Two Assumptions:

1. Three types of weather: sunny, rainy & foggy
2. The weather lasts all day.
Weather Prediction

We collect following probabilities:

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

For example:

\[ P(w_4 = \text{Rainy} \mid w_3 = \text{Foggy}, w_2 = \text{Sunny}, w_1 = \text{Sunny}) \]
A problem:

the larger $n$ is, the more statistics we must collect.

When $n = 5$, $3**5 = 243$ past histories!
Weather Prediction

Markov assumption:

In a sequence \{w_1, w_2, ..., w_n\}:

\[
P(w_n \mid w_{n-1}, w_{n-2}, \ldots, w_1) \approx P(w_n \mid w_{n-1})
\]

First-order assumption
Weather Prediction

Joint probability of $P(w_1, w_2, \ldots, w_n)$:

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

Only $3^2 = 9$ probability cases!
### Weather Prediction

Some numbers for $P(w_{\text{tomorrow}}|w_{\text{today}})$:

<table>
<thead>
<tr>
<th>Today’s Weather</th>
<th>Tomorrow’s Weather</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sunny</td>
</tr>
<tr>
<td>Sunny</td>
<td>0.8</td>
</tr>
<tr>
<td>Rainy</td>
<td>0.2</td>
</tr>
<tr>
<td>Foggy</td>
<td>0.2</td>
</tr>
</tbody>
</table>
Automaton:
An Example

Today is sunny, what’s the probability that tomorrow is sunny and the day after is rainy?

\[
P(w_2 = \text{Sunny}, w_3 = \text{Rainy} | w_1 = \text{Sunny}) = P(w_3 = \text{Rainy} | w_2 = \text{Sunny}, w_1 = \text{Sunny}) \times P(w_2 = \text{Sunny} | w_1 = \text{Sunny})
\]

\[
P(w_3 = \text{Rainy} | w_2 = \text{Sunny})
\]

\[
P(w_2 = \text{Sunny} | w_1 = \text{Sunny})
\]

\[
= P(w_3 = \text{Rainy} | w_2 = \text{Sunny}) \times P(w_2 = \text{Sunny} | w_1 = \text{Sunny})
\]

\[
= (0.05)(0.8)
\]

\[
= 0.04
\]
Another Example

Today is foggy, what’s the probability that it will be rainy two days from now?

\[
P(w_3 = \text{Rainy} \mid w_1 = \text{Foggy}) = P(w_2 = \text{Foggy}, w_3 = \text{Rainy} \mid w_1 = \text{Foggy}) + \\
P(w_2 = \text{Rainy}, w_3 = \text{Rainy} \mid w_1 = \text{Foggy}) + \\
P(w_2 = \text{Sunny}, w_3 = \text{Rainy} \mid w_1 = \text{Foggy}) + \\
= P(w_3 = \text{Rainy} \mid w_2 = \text{Foggy})P(w_2 = \text{Foggy} \mid w_1 = \text{Foggy}) + \\
P(w_3 = \text{Rainy} \mid w_2 = \text{Rainy})P(w_2 = \text{Rainy} \mid w_1 = \text{Foggy}) + \\
P(w_3 = \text{Rainy} \mid w_2 = \text{Sunny})P(w_2 = \text{Sunny} \mid w_1 = \text{Foggy}) \\
= (0.3)(0.5) + (0.6)(0.3) + (0.05)(0.2) \\
= 0.34
\]
You are locked in a room and asked the weather outside. The only evidence is whether the person carrying the meal is carrying an umbrella or not.

<table>
<thead>
<tr>
<th></th>
<th>Probability of Umbrella</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunny</td>
<td>0.1</td>
</tr>
<tr>
<td>Rainy</td>
<td>0.8</td>
</tr>
<tr>
<td>Foggy</td>
<td>0.3</td>
</tr>
</tbody>
</table>
Based on Bayes’ Rule:

\[
P(w_1, \ldots, w_n \mid u_1, \ldots, u_n) = \frac{P(u_1, \ldots, u_n \mid w_1, \ldots, w_n) P(w_1, \ldots, w_n)}{P(u_1, \ldots, u_n)}
\]

\[
P(w_1, \ldots, w_n) = \prod_{i=1}^{n} P(w_i \mid w_{i-1})
\]

\[
P(u_1, \ldots, u_n \mid w_1, \ldots, w_n) = \prod_{i=1}^{n} P(u_i \mid w_i)
\]

\[
P(u_1, \ldots, u_n) \text{ is prior probability of seeing a particular sequence of umbrella events}
\]
Example 1

The day you were locked in is sunny.
The next day, caretaker carried an umbrella.
The prior probability of caretaker carrying an umbrella is 0.5
What’s the probability that the second day was rainy?
Example 1

\[
\begin{align*}
P(w_2 = \text{Rainy} | w_1 = \text{Sunny}, u_2 = \text{True}) &= \frac{P(w_2 = \text{Rainy}, w_1 = \text{Sunny} | u_2 = \text{T})}{P(w_1 = \text{Sunny} | u_2 = \text{T})} \\
(u_2 \text{ and } w_1 \text{ independent}) &= \frac{P(w_2 = \text{Rainy}, w_1 = \text{Sunny} | u_2 = \text{T})}{P(w_1 = \text{Sunny})} \\
(Bayes' Rule) &= \frac{P(u_2 = \text{T} | w_1 = \text{Sunny}, w_2 = \text{Rainy})P(w_2 = \text{Rainy}, w_1 = \text{Sunny})}{P(w_1 = \text{Sunny})P(u_2 = \text{T})} \\
(Markov \ assumption) &= \frac{P(u_2 = \text{T} | w_2 = \text{Rainy})P(w_2 = \text{Rainy}, w_1 = \text{Sunny})}{P(w_1 = \text{Sunny})P(u_2 = \text{T})} \\
(P(A, B) = P(A | B)P(B)) &= \frac{P(u_2 = \text{T} | w_2 = \text{Rainy})P(w_2 = \text{Rainy} | w_1 = \text{Sunny})P(w_1 = \text{Sunny})}{P(w_1 = \text{Sunny})P(u_2 = \text{T})} \\
(Cancel : P(\text{Sunny})) &= \frac{P(u_2 = \text{T} | w_2 = \text{Rainy})P(w_2 = \text{Rainy} | w_1 = \text{Sunny})}{P(u_2 = \text{T})} \\
&= \frac{(0.8)(0.05)}{0.5} \\
&= .08
\end{align*}
\]
Example 2

The day you were locked in is sunny. Caretaker carried an umbrella on day 2 but not day 3.
The prior probability of caretaker carrying an umbrella is 0.5
What's the probability that the day 3 was foggy?
Example 2

\[
P(w_3 = F | \ w_1 = S, u_2 = T, u_3 = F) = P(w_2 = \text{Foggy}, w_3 = \text{Foggy} | \ w_1 = \text{Sunny}, u_2 = \text{True}, u_3 = \text{False}) + P(w_2 = \text{Rainy}, w_3 = \text{Foggy} | \ldots) + P(w_2 = \text{Sunny}, w_3 = \text{Foggy} | \ldots)
\]

\[
P(u_3 = F | w_3 = F)P(u_2 = T | w_2 = F)P(w_3 = F | w_2 = F)P(w_2 = F | w_1 = S)P(w_1 = S)
\]

\[
P(u_3 = F | w_3 = F)P(u_2 = T | w_2 = R)P(w_3 = F | w_2 = R)P(w_2 = R | w_1 = S)P(w_1 = S)
\]

\[
P(u_3 = F | w_3 = F)P(u_2 = T | w_2 = S)P(w_3 = F | w_2 = S)P(w_2 = S | w_1 = S)P(w_1 = S)
\]

\[
P(u_3 = F | w_3 = F)P(u_2 = T | w_2 = F)P(w_3 = F | w_2 = F)P(w_2 = F | w_1 = S)
\]

\[
P(u_3 = F | w_3 = F)P(u_2 = T | w_2 = R)P(w_3 = F | w_2 = R)P(w_2 = R | w_1 = S)
\]

\[
P(u_3 = F | w_3 = F)P(u_2 = T | w_2 = S)P(w_3 = F | w_2 = S)P(w_2 = S | w_1 = S)
\]

\[
P(u_3 = F)P(u_2 = T)
\]

\[
\frac{(0.7)(0.3)(0.5)(0.15)}{(0.5)(0.5)} + \frac{(0.7)(0.8)(0.2)(0.05)}{(0.5)(0.5)} + \frac{(0.7)(0.1)(0.15)(0.8)}{(0.5)(0.5)} = 0.119
\]
Relationship to Speech

\[ \arg\max_{w \in \mathcal{L}} P(w | y) \]

W is a string of words
L is the language you are interested in
Y is the set of acoustic vectors
Based on Bayes’ Rule:

\[
\text{argmax}_{w \in \mathcal{L}} P(w|y) = \text{argmax}_{w \in \mathcal{L}} \frac{P(y|w)P(w)}{P(y)}
\]

For a single speech input, \(P(y)\) will be constant.

\[
\text{argmax}_{w \in \mathcal{L}} P(y|w)P(w)
\]
Word Pronunciations

Calculate $P(y|w)$ ---- Use phone

Take “of” as an example:

\[
P(y|w_{of}) = P(y|Start_{ax_{-}ah_{-}v_{-}End}) + P(y|Start_{ah_{-}v_{-}End})
\]

\[
P(y|w_{of}) = P(q_b|q_a)P(y_0|q_b)P(q_c|q_b)P(y_1|q_c) + P(q_d|q_a)P(y_0|q_d)P(q_c|q_d)P(y_1|q_c)
\]
\( P(w) \) is the prior probability of the string of words

\[
P(w) = \prod_{i=1}^{n} P(w_i | w_{i-1})
\]

"I like cabbages"

\[
P(\text{I like cabbages}) = P(\text{I}|\text{START})P(\text{like}|\text{I})P(\text{cabbages}|\text{like})
\]

Trigram grammar

\[
P(w_n | w_{n-1}, w_{n-2})
\]
Acknowledge

THANK YOU