Hidden Markov Chain - HMM

Miao Cai

4/16/2019

1 Stackoverflow answers

Two answers from stackoverflow

1. by matt

To explain by example, I'll use an example from natural language processing. Imagine you want to know the probability of this sentence:

I enjoy coffee

In a Markov model, you could estimate its probability by calculating:

`P(WORD = I) x P(WORD = enjoy | PREVIOUS_WORD = I) x P(word = coffee| PREVIOUS_WORD = enjoy)` Now, imagine we wanted to know the parts-of-speech tags of this sentence, that is, if a word is a past tense verb, a noun, etc.

We did not observe any parts-of-speech tags in that sentence, but we assume they are there. Thus, we calculate what's the probability of the parts-of-speech tag sequence. In our case, the actual sequence is:

> PRP-VBP-NN

(where PRP="Personal Pronoun", VBP="Verb, non-3rd person singular present", NN="Noun, singular or mass". See https://cs.nyu.edu/grishman/jet/guide/PennPOS.html for complete notation of Penn POS tagging)

But wait! This is a sequence that we can apply a Markov model to. But we call it hidden, since the parts-of-speech sequence is never directly observed. Of course in practice, we will calculate many such sequences and we'd like to find the hidden sequence that best explains our observation (e.g. we are more likely to see words such as 'the', 'this', generated from the determiner (DET) tag)

The best explanation I have ever encountered is in a paper from 1989 by Lawrence R. Rabiner: http://www.cs.ubc.ca/~murphyk/Bayes/rabiner.pdf

2. by Aerin

Since Matt used parts-of-speech tags as an HMM example, I could add one more example: Speech Recognition. Almost all large vocabulary continuous speech recognition (LVCSR) systems are based on HMMs.

"Matt's example": **I enjoy coffee**

In a **Markov model**, you could estimate its probability by calculating:

P(WORD = I) x P(WORD = enjoy | PREVIOUS_WORD = I) x P(word = coffee| PREVIOUS_WORD = enjoy) In a Hidden Markov Model,

Let's say 30 different people read the sentence "I enjoy hugging" and we have to recognize it. Every person will pronounce this sentence differently. So we do NOT know whether or not the person meant "hugging" or "hogging". We will only have the probabilistic distribution of the actual word.

In short, a **hidden Markov model** is a statistical Markov model **in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states**.

2 Theories

2.1 Define an HMM

3 parts: π, A, and B.

1. π: probability of starting at state i,
2. $A(i, j)$: probability of going to state j from state i,
3. $B(j, k)$: probability of observing symbol k in state j.

More than just just Markov assumptions.

With Markov models, we need to have two things: getting the probability of a sequence and train the model. But traing the models here will be harder with the limits of numerical accuracy of float. We also have one more task: get the most-likely hidden state sequence.

2.2 Gaussian Mixture Models + HMMs

Choose a state -> choose a Gaussian -> generate a value from the Gaussian

3 R examples

- https://blog.revolutionanalytics.com/2014/03/r-and-hidden-markov-models.html
- https://inovancetech.com/hmm-tutorial-1.html
- https://www.r-bloggers.com/hmm-example-with-depmixs4/