An NLP Tip

Javad.Pourmostafa

October 2019

1 HMM-Viterbi Algorithm Example

Consider the following rhyme from "Jabberwocky" by Lewis Caroll:

'Twas brillig, and the slithy toves
Did gyre and gimble in the wabe;
All mimsy were the borogoves,
And the mome raths outgrabe.

Let's take the last phrase,**"the mome raths outgrabe"**, as an example. Let's build a hidden Markov model for predicting part of speech tags in this sentence. For simplicity, suppose that we have just three tags: **N (noun), V (verb) and D (determiner).** We need to specify initial probabilities of these tags and transition probabilities from one tag to another. Surely, these probabilities may be estimated using some annotated corpus. But let's suppose for now that all tags and all transitions are equiprobable:

$P(N|start) = p(O|start) = p(D|start) = \frac{1}{3}$

$P(N|N) = p(V|N) = p(D|N) = \frac{1}{3}$

and so on.

Dealing with HMMs, we also need to specify output probabilities of words given the tag. For simplicity, we consider the following outcomes:

$N : mome|raths|outgrabe$

$V : raths$

$D : the|a$

Let all these outcomes be also equiprobable, i. e,

$p(mome|N) = p(raths|N) = p(outgrabe|N) = 1/3$

$p(raths|V) = 1$

$p(the|D) = p(a|D) = 1/2$

Given this toy model, let's find the probabilities of possible tag sequences for the phrase **"the mome raths outgrabe"**. In other words, these are the conditional probabilities: $p(XXXX|phrase)$, where each tag X is either N, or V, or D.

First question for you: how many different tag sequences exist?

Second question: which of them could happen in our case with the transition and output probabilities defined above?

Answers: there are $3^4 = 81$ sequences, but only **two of them** are possible in our case. "the" can be generated only from D, "mome" and "outgrabe" can be generated only from N, and "raths" can be generated wither from N or V. So we can have either **DNNN** or **DNVN**.

So we have just seen, that probabilities of **79** tag sequences are equal to 0, and we need to compute these two: $p(DNNN|phrase)$ and $p(DNVN|phrase)$. According to the HMM model, the joint probabilities are:

$p(DNVN, phrase) = p(D|start)p(the|D)p(N|D)p(mome|N)p(V|N)p(raths|V)p(N|V)p(outgrabe|N)$

$p(DNNN, phrase) = p(D|start)p(the|D)p(N|D)p(mome|N)p(N|N)p(raths|N)p(N|N)p(outgrabe|N)$

$p(DNNN|phrase) = x \times 1/3$

$p(DNVN|phrase) = x \times 1$