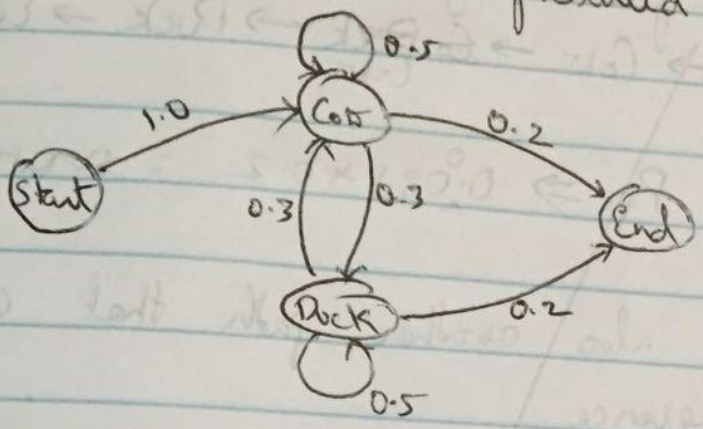
NLP - Assignment 2

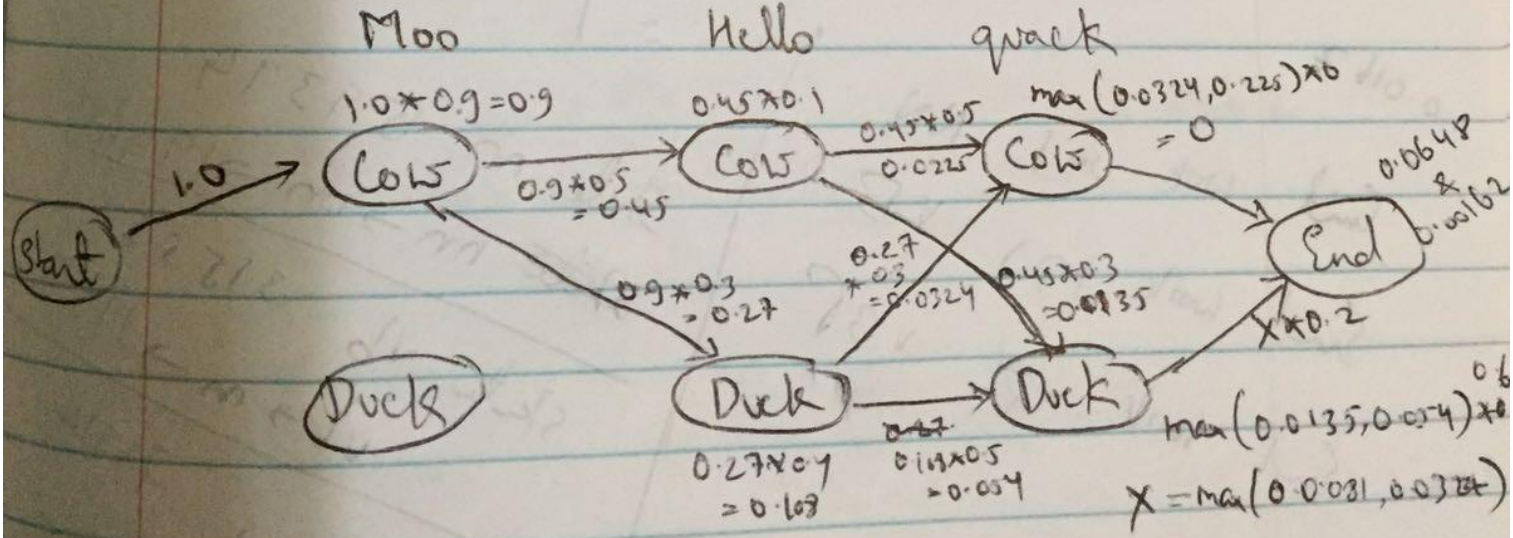
i) The states can be represented as



Cow can emits 'moo' (with 0.9 probability) or
'hello' (with 0.1 probability)

Duck emits 'quack' (0.6 probability) or
'hello' (0.4 probability)

Using viterbi algorithm, the states can be as follows:



Looking at the diagram, the path that emits maximum probability is

Start → Cow → Duck → Duck → End

$P_1 = 0.0324 * 0.2 = 0.00648$.

b) There is another path that can emit the same sequence

Start → Cow → Cow → Duck → End

$P_2 = 0.00081 * 0.2 = 0.000162$.

Total probability of emitting this sentence is

$P_1 + P_2 = 0.00648 + 0.000162$

$= 0.00081$

2) Correct Sentence

I want to work

I : PRP
want : VBP
to : TO
work : VB

Incorrect Sentence

Engineers work day and night.

Engineers : NNS
work : ~~VBP~~ NN
day : NN
and : CC
night : NN

Here 'work' should have been verb.

~~let's look at learning~~

The relative probability that ~~the~~ tagger tagged it ~~was~~ wrong is.

Relative probability that is of correct tagged and wrong tagged.

$$\frac{P(\text{Wrong tagged})}{P(\text{Correct tagged})} = \frac{P(\text{Work as NN}) * P(NN \mid NNS) * P(NN \mid NN)}{P(\text{Work as verb}) * P(VB \mid NNS) * P(NN \mid VB)}$$

(NN ← previous state)

From Corpus,

(1) $P(\text{Work as NN}) = \frac{301}{458} = 0.657$

(2) $P(\text{Work as VB}) = \frac{131}{458} = 0.286$

(3) state NN = 159394

NN → NN = 19463

$\Rightarrow P(NN \mid NN) = \frac{19463}{159394} = 0.122$

(4) state VB = 31555

VB → NN = 1957

$P(NN \mid VB) = \frac{1957}{31555} = 0.062$

(5) state NNS = 71914

NNS → VB = 285

$$P(VB|NNS) = \frac{285}{71914} = 0.003963$$

NNS → NN = 1517

$$P(NN|NNS) = \frac{1517}{71914} = 0.02109$$

Note that, in relative probability, we are interested in terms that differ in value.

$$\frac{P(\text{Wrong tagged})}{P(\text{Correct tagged})} = \frac{0.657 * 0.02109 * 0.122}{0.286 * 0.003963 * 0.062}$$

$$= \frac{0.00169}{0.0000702}$$

Note that P(Wrong tagged) has higher value compared to P(Correct tagged). That is why the tagger tagged it wrong.