

Q2. JET HMM Tagger:

a) The sentence "The investment failed" is correctly tagged.

-----  
Sentence: The investment failed.  
Annotate The == <constit cat="DT">  
Annotate investment == <constit cat="NN">  
Annotate failed == <constit cat="VBD">  
Annotate . == <constit cat=".">

b) In sentence "The investment is failed", every word is tagged correctly, except "failed". It is supposed to be in the category 'JJ'.

-----  
Sentence: The investment is failed.  
Annotate The == <constit cat="DT">  
Annotate investment == <constit cat="NN">  
Annotate is == <constit cat="VBZ">  
Annotate failed == <constit cat="VBN">  
Annotate . == <constit cat=".">

**The investment is failed.**

**DT NN VBZ VBN .**

STATE start 47396  
ARC TO DT 9748  
STATE DT 98334  
ARC TO NN 46646  
STATE NN 159394  
ARC TO VBZ 6967  
STATE VBZ 25735  
ARC TO VBN 3817  
STATE VBN 24107  
EMIT failed 53  
ARC TO . 1202  
STATE . 47396  
ARC TO end 47396  
State end

The computation of relative probability:

$E(\text{'failed' | VBN}) * P(\text{DT | start}) * P(\text{NN | DT}) * P(\text{VBZ | NN}) * P(\text{VBN | VBZ}) * P(\text{. | VBN}) =$   
 $(53 | 24107) * (9748 | 47396) * (46646 | 98334) * (6967 | 159394) * (3817 | 25735) * (1202 | 24107)$

**The investment is failed.**

**DT NN VBZ JJ .**

STATE start	47396
ARC TO DT	9748
STATE DT	98334
ARC TO NN	46646
STATE NN	159394
ARC TO VBZ	6967
STATE VBZ	25735
ARC TO JJ	1903
STATE JJ	73230
EMIT failed	6
ARC TO .	1780
STATE .	47396
ARC TO end	47396
State end	

The computation of relative probability:

$$\begin{aligned} &E(\text{'failed' | JJ}) * P(\text{DT | start}) * P(\text{NN | DT}) * P(\text{VBZ | NN}) * P(\text{JJ | VBZ}) * P(. | \text{JJ}) = \\ &(6 | 73230) * (9748 | 47396) * (46646 | 98334) * (6967 | 159394) * (1903 | 25735) * \\ &(1780 | 73230) \end{aligned}$$

From STATE start to STATE VBZ are common factor in two results, I just ignore their calculation. Thus, in short, I have:

$$\begin{aligned} &E(\text{'failed' | VBN}) * P(\text{VBN | VBZ}) * P(. | \text{VBN}) \\ &= (53 | 24107) * (3817 | 25735) * (1202 | 24107) \\ &\text{is greater than} \\ &E(\text{'failed' | JJ}) * P(\text{JJ | VBZ}) * P(. | \text{JJ}) \\ &= (6 | 73230) * (1903 | 25735) * (1780 | 73230) \end{aligned}$$

The relative probability of the right tagger('JJ') is much lower than the wrong tagger('VBN').

This leads to the wrong tagged result.