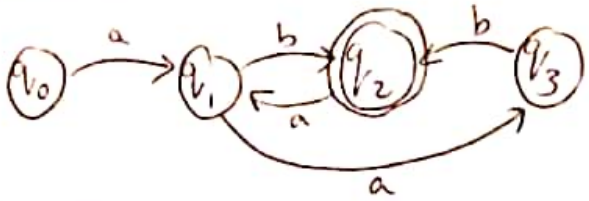
NLP HW2

B. Problems:

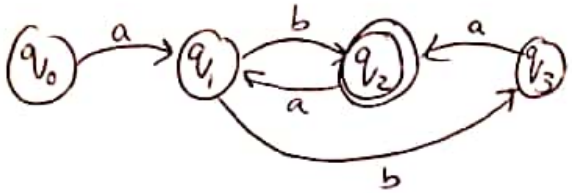
1. a.



Accepted strings = ab, abab, aab, aaba b, ...

$$RE = (ab \mid aab)(ab \mid aab)^*$$

1.b.



Accepted strings = ab, aba, abab, abaab, ...

$$RE = (ab \mid aba)(ab \mid aba)^*$$

2. Bigram Probabilities

Input Sentence (S) = "The standard Turbo engine is hard to work"

i) $P(S) = P(\text{The}) \times P(\text{standard} \mid \text{The}) \times P(\text{Turbo} \mid \text{standard}) \times P(\text{engine} \mid \text{turbo}) \times P(\text{is} \mid \text{engine}) \times P(\text{hard} \mid \text{is}) \times P(\text{to} \mid \text{hard}) \times P(\text{work} \mid \text{to})$

$$= \frac{3676}{68737} \times \frac{3}{3676} \times \frac{2}{10} \times \frac{0}{2} \times \frac{0}{17} \times \frac{0}{447} \times \frac{3}{4} \times \frac{7}{1551}$$

No smoothing = 0

Scanned with CamScanner

ii) Laplace smoothing (Add one)

$$P(\text{Sentence}) = P(\text{The}) \times P(\text{standard} \mid \text{The}) \times P(\text{Turbo} \mid \text{standard}) \times P(\text{engine} \mid \text{Turbo}) \times P(\text{is} \mid \text{engine}) \times P(\text{hard} \mid \text{is}) \times P(\text{to} \mid \text{hard}) \times P(\text{work} \mid \text{to})$$

$$= \frac{3676}{68737} \times \frac{(3+1)}{(3676+7602)} \times \frac{(2+1)}{(10+7602)} \times \frac{(0+1)}{(2+7602)} \times \frac{(0+1)}{(17+7602)} \times \frac{(0+1)}{(447+7602)} \times \frac{(3+1)}{(4+7602)} \times \frac{(7+1)}{(1551+7602)}$$

$$= 7.368E-27$$

iii) Good Turing Smoothing

$$P(\text{Sentence}) = P(\text{The}) \times P(\text{standard} \mid \text{The}) \times P(\text{Turbo} \mid \text{standard}) \times P(\text{engine} \mid \text{Turbo}) \times P(\text{is} \mid \text{engine}) \times P(\text{hard} \mid \text{is}) \times P(\text{to} \mid \text{hard}) \times P(\text{work} \mid \text{to})$$

$$= \frac{3676}{68737} \times \frac{(3+1) \times \frac{1363}{1813}}{66517} \times \frac{(2+1) \times \frac{1813}{6376}}{66517} \times \frac{16520}{66517} \times \frac{16520}{66517} \times \frac{16520}{66517} \times \frac{(3+1) \times \frac{1363}{1813}}{66517} \times \frac{(7+1) \times \frac{184}{205}}{66517}$$

$$= 2.318E-21$$

Scanned with CamScanner