Assignment 4
Advanced Natural Language Processing
Fall 2016

Anusha Ramamurthy
anuramam@iu.edu

October 4, 2016

Solutions

1. **FSA**

   (a) We define the transition matrix for FSA as given below:

$$\text{a} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\text{b} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

$$\text{c} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\text{init} = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}$$

$$\text{final} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

   (b) From the transition matrix we define the regular expression for the language given by FSA as

$$\lambda|(bb^*)|(a^+bb^*)|ca^*|(b^+a^+bb^*)|(aa^*bb^*)(aa^*bb^*)?$$

2. **Markov chains**

   (a) Thus the probability that we have 4 steps and exactly 3 lines are busy is given by

$$vP^4 = vPPPP = \left[\begin{array}{cccc} 0.18791 & 0.33393 & 0.27332 & 0.20484 \end{array} \right]$$

$$s(\xi_4 = s_3) = \mathbf{0.20484}$$

(b) Thus Line 1 has the highest probability of being busy.

3. **POS Tagging**
One of the methods to find the most likely sequence is by iterating over all combinations of ”POS Tags” and to select the one with the highest probability. I have used the log likelihood to get rid of the underflow. I found the below to be the sequence with the highest probability. Please find the code attached.

(’NN’, ’VBZ-HL’, ’IN’, ’NIL’, ’NN’)

4. **Read Ranaparkhi (1996)**

The author discusses the how to assign POS tags to unseen or unannotated words. He suggest the Maximum entropy method. In this method he suggests applying a value feature to find the entropy of each historis. This feature is defined as 0 or 1 based on the context of the historis. He suggests assigning rules like if the word ends with ”ing” and the tag is ”VB”. Similar rules can be defined to construct the features. The features now contribute to the probability of the historis. He suggests assigning a new tag to the unknown word and computing the entropy value for the entire sequence by looking at the tags assigned to the word immediately before and after the current word for which we are to determine the tag, and also by looking at the words two before and after. He also suggests doing a beam search to identify all the possible tags that can be assigned to the unknown word. This approach seems to be a good one as it takes in the context of the dataset as well as how the author of the dataset has a way of using the words.
The entropy is given by

$$H(p) = -\sum_{h\epsilon H, t\epsilon T} p(h,t)\log p(h,t)$$

and the expectation of features are given by:

$$Ef_j = \sum_{h\epsilon H, t\epsilon T} p(h,t) f_j(h,t)$$

Conclusion: While there appear to be other models that perform nearly as well, the maximum entropy model doesnt make assumptions about the probability distribution of the dataset. Hence it makes realistic determination of the probability and entropy values. Thus giving better accuracy over other methods.