Atieh Sarvi

Iran University of Science and Tech.
www.iust.ac.ir

Assignment 3 Problems

NLP : Fall 1400 : Dr. Minaei
Due Friday, Dey 3, 1400

Contents

Problem 1

In this section, you should do POS tagging using HMM method on this sentence: Can Tom mark watch? according to these four sentences:

Mark can watch
Will can mark watch
Can Tom watch?
Tom will mark watch

Assumption: we have three tags: verb, noun, modal.

(a)

First, apply POS tagging on the given sentences and create a table that shows the probability that each word is a noun, modal, or verb.
A sample table:

كلمات	noun	modal	verb
Tom	2\6	0	0

(b)

In the next step, two more tags should be added to the first and last sentences. $< S >$ is added to the beginning of the sentence and $< E >$ to the end of the sentence.

<S> Mark can watch<E>
<S> Will can mark watch<E>
<S> Can Tom watch?<E>
<S> Tom will mark watch<E>

Now draw a table and calculate the probability of two labels occurring together.
Sample table:

	Noun	Modal	Verb	<E>
<S>				
Noun				
Modal				
Verb				

(c)

In this section, according to the calculated probabilities, perform POS tagging for this sentence:
can Tom mark watch?
(Draw a graph of the sentence, then delete the edges that have a probability of zero to get the answer.)
NOTE: for more information check this link.
$https://www.mygreatlearning.com/blog/pos-tagging/$

Problem 2

According to the following grammar and using the CKY algorithm, perform the parsing operation by drawing a table for the following sentence:

- Astronomers saw stars with ears

S → NP VP	1.0	NP → NP PP	0.4
PP → P NP	1.0	NP → astronomers	0.1
VP → V NP	0.7	NP → ears	0.18
VP → VP PP	0.3	NP → saw	0.04
P → with	1.0	NP → stars	0.18
V → saw	1.0	NP → telescopes	0.1

astronomers   saw   stars   with   ears

Problem 3

Suppose you want to find all of Freud's psychology textbooks that contain the words "satisfaction," "goals," and "rich" without the words "harassment" and "hate."

(a)

What is the simplest technique that comes to your mind to find these books?

(b)

Do you think your proposed method works well in practice? Discuss possible problems.

(c)

As it is clear, the data in this issue is unstructured. Can you find a way to turn it into a structured one? Explain.

(d)

Based on your answer to the previous section, provide a way to find the books you are looking for.

Problem 4

(a)

Explain the limitations of PCFGs using an example.

(b)

How does lexicalized grammar reduce these limitations? Explain with a simple example.

Problem 5

Suppose you work in a medical research center and want to use your NLP knowledge to take a step in the development of the medical community.

(a)

What can be extracted from medical texts using NER?

(b)

Suppose you want to design a system that extracts these items from medical texts. Describe the design, development, and evaluation steps of this system in detail.
NOTE: If you answer this question in detail and develop your ideas in a decent way, you might get some extra points!

Notes

- If you have any questions, feel free to ask. You can ask your questions in the Telegram group.

- Please upload your assignments as a zipped folder with all necessary components. Upload your file in $HW3\_NLP\_Y ourStudentID\_Y ourName.zip$ format.