Assignment 2

October 31, 2016 (Deadline: November 14, 2016)

IEMS Assignment 2: Yelp comments sentiment analysis

1 Problem

Your task is to classify the sentiments expressed in the comments of Yelp (American restaurant review, similar to Openrice in HK). The comments are in form of a complete sentence, taken from the reviews of restaurants in the Arizona state. You should classify the sentiments as positive, neutral or negative. Negation (not / never / no etc.) should be handled.

2 Tools

You may use the WordNet (http://sentiwordnet.isti.cnr.it/) or its processed version (wordcsv_simp) in tutorial 3 as the corpus. NLTK should be used in this assignment, but the other .py files in Tutorial 3 are NOT allowed to be used. Hint: you should analyze the POS of the words.

3 Input

We have one file **test.txt**. The **test.txt** file stores 15 testing sentences. The first 5 sentences are positive comments. Next 5 are neutral, and the last 5 are negative.

4 Submission

A runnable Python program file should be submitted (assignment2.py). The output txt (result.txt) should be submitted as well. It is in the form of original sentence, follow by a white space and the annotation. Here is the annotation for different sentiments:

positive	+
neutral	0
negative	-

Sample lines of result.txt:

I don't feel comfortable with this restaurant. -
I am happy with the salsa here. +

5 Output

You need to print the accuracy of your programme with review comments. After runing the python program, the screen should display: (actual number varies)

```
>>>classifier's accuracy on review comments:  0.66
```

6 Grading Scheme

50%	Complete runnable Python program
20%	Correct format of **result.txt** and program output
30%	Accuracy of the sentiment analysis

$$Accuracy = \frac{no.\ of\ correctly\ classified\ sentences}{total\ no.\ of\ test\ sentences\ (=15)}$$

The scoring of the accuracy part is relative. The submission with the best performance gets full mark. Others get a relative proportion of the mark.