Intellimind Assignment 1

Sentiment Aanalysis

- Preprocessing Of Data

- To preprocess the data using python i have imported re library and imported stopword from nltk.corpus.

  Done the Preprocessing and sentiment Analysis using word splitting.

- I have compile my text document using re.compile() function by making use of regular expression;
- I have used stopword library function to to remove unneccesary character from my document and saved this to my temporary file to do rest of my proccessing ;
- Then I have Compiled the code using regular expression of digits[0-9], Non character values using "\W", and repeat((a+b)+ followed by dot) on my temporary file to remove all useless character i have in my original text document ;
- I have Also used .replace() function to further preprocess my text data;
- All the preprocessd data are than saved in the form of the python list(contents{user define list(contain usefull contents which is then used for sentiment analysis infurther steps's)});

-Generate Sentiment Score For Each Documents

- Now, for sentiment analysis i have make use of the sentiwordnet , imported in python using nltk.corpus.
- Now for each word in my contents list i determine my positive, negative and objectivity score and then calculated the average to get the Score for my document.
- This I have done for each document.
- Here is my little summary for neural word, positive word and negative words.

Neural words are words with objectivity equal to 0 i.e.,if positive score and negative score is zero.

if negative score and positive score are equal to 0.5 than our objectivity is zero we also refer to them as neural, As the direct consequences of thse values for the positive and negative score, the objectivity score of these word is zero, despite there "neural" label. We will name them as half positive and half Negative words.

We also have weak positive and weak negative words for them the value of objective score varies from 0.166 to 0.98.

- Now using this i have determined the number of positive and Negative Documents From the list of the documents used.

Named Entity Recognition

- Imported spacy to be used for figure of speech.
- Used en_core_web_sm to for importing english language.
- And, then perform the named-entity recognition within text data for each Documents given to us.

Test Cases

- I have used two text file as my Test Cases one is of 10kb and Another of 1000kb.
- Results for my test cases, Format:[positive score, negative score, objectivity]

[0.0125, 0.10625, 0.88125]
[0.0250, 0.11875, 0.85625]

- Same Text documents are used as my Test cases for Named Entity Recognition