CSCE 638: Natural Language Processing

Assignment 2 Report

Submitted By – Sameer Kumar Behera UIN: 526004296

System Requirements:

- Python must be installed
- Installation Link: https://www.python.org/downloads/
- Pandas Library in Python should be installed too.

Compile and Run Method:

1. Open a Terminal

2. Go to the Project Folder, i.e. PA2-638/python

3. On the Terminal, Enter the Following Commands For:

- Naive Bayes Classifier
  - python NaiveBayes.py ../data/imdb1
  - python NaiveBayes.py -f ../data/imdb1
  - python NaiveBayes.py -b ../data/imdb1
- Perceptron
  - python Perceptron.py ../data/imdb1/ 1
  - python Perceptron.py ../data/imdb1/ 10
  - python Perceptron.py ../data/imdb1/ 50
  - python Perceptron.py ../data/imdb1/ 100
  - python Perceptron.py ../data/imdb1/ 500

Result and Analysis:

REMARK – The Codes were run on Ubuntu and Cloud9 Platform and the Accuracy Results slightly varied on both systems.

- **Naive Bayes Classifier**
  - python NaiveBayes.py ../data/imdb1
    - Average Accuracy = 81.65%
  - python NaiveBayes.py -f ../data/imdb1
    - Average Accuracy = 81.10%
  - python NaiveBayes.py -b ../data/imdb1
    - Average Accuracy = 82.90%

Binarized version of Naive Bayes which relies on occurrence of word in a document rather than frequency, performs slightly better than general Naïve Bayes. Removal of stop words doesn't help in improving the accuracy.

- **Perceptron**
  - python Perceptron.py ../data/imdb1/ 1
    - Average Accuracy = 50%
  - python Perceptron.py ../data/imdb1/ 10
    - Average Accuracy = 50.7%
  - python Perceptron.py ../data/imdb1/ 50
    - Average Accuracy = 69.1%
  - python Perceptron.py ../data/imdb1/ 100
    - Average Accuracy = 81.4%

- python Perceptron.py ../data/imdb1/ 500
  - Average Accuracy = 82.8%

Clearly, with increase in the iterations, the perceptron classification accuracy performance increases tremendously.

Any Bugs/Limitations:

No such bugs/limitations as per my knowledge.