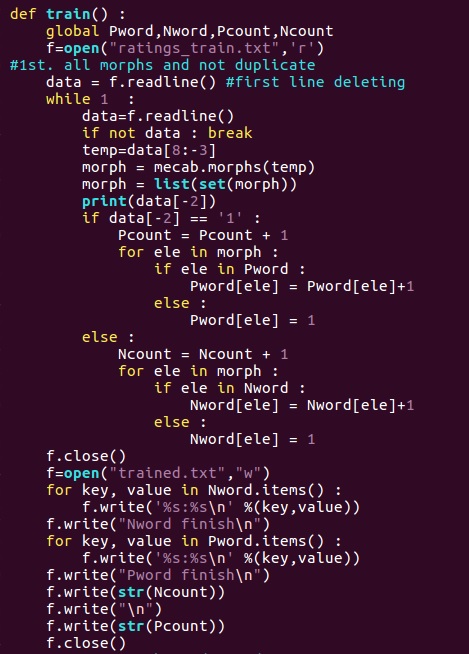
Assignment2

2013011491

컴퓨터전공

안찬영

기본적으로 konlpy와 math module을 import하였으며, 결과적으로는 konlpy.tag에서도 Mecab관련 api만 사용하였습니다.



Train 함수입니다. “ratings\_train.txt” 파일을 한 줄 씩 입력 받아 mecab을 이용해 형태소를 나누고 label이 긍정일 때는 Pword 딕셔너리에 해당 형태소를 key로 추가합니다. 혹은 이미 해당 키에 대해 딕셔너리 키가 존재할 경우 value값을 하나 증가시킵니다. Label이 부정일 때는 Nword 딕셔너리에 똑같이 해줍니다. 그 후 ‘trained.txt’ 파일을 생성해 Nword, Pword 딕셔너리의 key와 value를 저장하고, label에서 긍정 count와 부정 count도 저장합니다.



main함수입니다.

첫번째 train()함수는 ‘trained.txt’파일을 생성 할 최초만 수행하고, 그 후에는 주석처리를 통해 막아 놓았습니다. ‘trained.txt’파일로부터 Nword와 Pword를 불러와서 test파일에서 예제들을 분류할 기준이 됩니다. 테스트할 파일을 열고 각각의 line 단위로 시행되며, 해당 line을 형태소 분석 한 뒤, naïve bayes 분류법을 이용해 각각이 긍정일 확률과 부정일 확률을 비교해 더 큰 확률을 결정한 뒤 “ratings\_result.txt”파일에 label 값을 추가해서 파일쓰기를 한 뒤 종료합니다.

결과 예시

