



Week 11

Text Analysis

May 3, 2021
Daeun Lee, Jeewoo Yoon
delee12@skku.edu

Assignment

- 11주차 실습 영상에서 배운 내용 (sentiment analysis- nltk SentiWordNet)을 활용하여, 간단한 감성분석 모델 만들고 리포트 및 코드 제출
- 데이터
 - polarity.csv (첨부)
- 제출물
 - 리포트 (.pdf)
 - 감성분석 성능평가 결과
 - 코드 파일 (.ipynb)
- 제출 기한 : 2021.5.9
- RTS QnA : 2021. 5. 12

Assignment

- 필요 columns
 - ngram : 한글 단어 추출
 - NEG : negative score
 - POS : positive score

ngram	freq	COMP	NEG	NEUT	POS	max.value	max.prop
가*/JKS	1	0	0	0	1	POS	1
가*/JKS;있/VV	1	0	0	0	1	POS	1
가*/JKS;있/VV;있/EP	1	0	0	0	1	POS	1
가*/VV	3	0	0	0	1	POS	1
가*/VV;ㄴ다*/EF	1	0	0	0	1	POS	1
가/JKC	17	0	0.411764706	0.117647059	0.470588235	POS	0.470588235
가/JKC;되/VV	11	0	0.272727273	0.181818182	0.545454545	POS	0.545454545
가/JKC;되/VV;ㄴ/ETM	2	0	0	1	0	NEUT	1
가/JKC;되/VV;ㄹ/ETM	1	0	0	0	1	POS	1
가/JKC;되/VV;어/EC	2	0	0.5	0	0.5	NEG	0.5
가/JKC;되/VV;어이지요/EF	1	0	0	0	1	POS	1
가/JKC;아니/VCN	6	0	0.666666667	0	0.333333333	NEG	0.666666667
가/JKC;아니/VCN;ㄴ가/EC	1	0	1	0	0	NEG	1
가/JKC;아니/VCN;면/EC	1	0	0	0	1	POS	1
가/JKC;아니/VCN;버니다/EF	1	0	0	0	1	POS	1
가/JKS	112	0	0.339285714	0.026785714	0.633928571	POS	0.633928571
가/JKS;가능/NNG	1	0	0	0	1	POS	1
가/JKS;가능/NNG;하/XSA	1	0	0	0	1	POS	1

Assignment

- To-do
- $\text{total sentiment score} = \text{positive score} - \text{negative score}$
- 첨부된 라라랜드 영화 리뷰 데이터를 활용하여 (`naver_preprocess.csv`), 각 문장마다 `positive`에 해당하는 단어, `negative`에 해당하는 단어의 점수를 활용하여 문장의 `sentiment score`를 구한다. (week11 text_analysis - nltk sentiwordnet 코드 참고)
- `sentiment score`와 실제 평가 `score`와 비교하여 성능평가 (정확도 점수 성적과 무관)
 - labeling:
 - 1~5 : negative (0)
 - 6~10 : positive (1)

Assignment

- 성능평가 코드 예시
 - y_true : 실제 라벨
 - y_pred : 예측 라벨

```
from sklearn.metrics import classification_report
y_true = [0, 1, 1, 0, 0]
y_pred = [0, 0, 1, 1, 0]
target_names = ['negative', 'positive']
print(classification_report(y_true, y_pred, target_names=target_names))
```

	precision	recall	f1-score	support
negative	0.67	0.67	0.67	3
positive	0.50	0.50	0.50	2
accuracy			0.60	5
macro avg	0.58	0.58	0.58	5
weighted avg	0.60	0.60	0.60	5