

# 한국어 감성분석기

201213304 강현구

# 개요

## 텍스트에서 사람의 감성을 분석

### ◆ SNU KOSAC – Polarity.csv

감성 빈도율이 50%이상인 단어들을 추려서 만들어놓은  
N그램 16000개 단어가 포함된 감성 사전

### ◆ NSMC(Naver sentiment movie corpus)

네이버 영화 리뷰를 별점에 따라서  
0~5점인경우 0, 6~10점인경우 1로 분류해놓은 코퍼스

변경전

문장



형태소 분석



Ngram



문장의 감성비율  
및 극성 계산

변경후

리뷰



형태소 분석  
Feature 생성



기계 학습



극성 예측

## 네이버 영화 감성 말뭉치

NSMC(Naver sentiment movie corpus)  
(<https://github.com/e9t/nsmc>)

영화 리뷰를 별점에 따라서 극성분류

0~5점 → 0

6~10점 → 1

Id	document	label
9730759	내 나이와 같은 영화를 지금 본 나는 감동적이다..하지만 훗날 다시보면대사하나하나 그 감정을완벽하게 이해할것만 같다...	1

document를 꼬꼬마 형태소 분석기로 형태소 분석 →

형태소 분석된 words들을 vectorize하여 corpus\_feature 생성하고 label 생성

## 개선 사항

Only POSITIVE / NEGATIVE

### 감성어 사전(SNU KOSAC) – Polarity.csv

감성어 사전에서 긍/부정 특성만 이용하여  
극성 판별을 위한 feature로 추가한다.

COMP	NEG	NEUT	None	POS	max.value	max.prop
0	0	0	0	1	POS	1
0	0	0	0	1	POS	1
0	0	0	0	1	POS	1
0	0	0	0	1	POS	1
0	0	0	0	1	POS	1
0	0.470588	0.235294	0	0.294118	NEG	0.470588
0	0.363636	0.272727	0	0.363636	NEG	0.363636
0	0	1	0	0	NEUT	1
0	0	1	0	0	NEUT	1
0	1	0	0	0	NEG	1
0	0	0	0	1	POS	1
0	0.666667	0.166667	0	0.166667	NEG	0.666667
0	1	0	0	0	NEG	1
0	0	0	0	1	POS	1
0	1	0	0	0	NEG	1
0.017857	0.464286	0.080357	0.026786	0.410714	NEG	0.464286

형태소 분석된 words들을 dictionary탐색 ➔ Vectorized corpus feature에 긍부정 feature append

ex) Positive - [0,0, ..... 0,0] + [0,1]  
Negative - [0,0, ..... 0,0] + [1,0]  
else - [0,0, ..... 0,0] + [0,0]

## 결과

2000개의 review data를 train(1500), test(500) 한다.

### Normal Vectorized feature

	precision	recall	f1-score	support
negative	0.72	0.74	0.73	244
positive	0.74	0.73	0.74	256
avg / total	0.73	0.73	0.73	500



### Added sentimental dictionary feature

#### Words count feature

	precision	recall	f1-score	support
negative	0.73	0.74	0.73	244
positive	0.75	0.73	0.74	256
avg / total	0.74	0.74	0.74	500

#### Words value feature

	precision	recall	f1-score	support
negative	0.72	0.75	0.73	244
positive	0.75	0.73	0.74	256
avg / total	0.74	0.74	0.74	500

감사합니다.