

파이썬을 통해 분석한 『젊은 베르테르의 슬픔』

- 베르테르의 감정을 중심으로

디지털인문학입문II 01분반 06조
발하임01

0. 주제 선정 이유와 사용 모델

<젊은 베르테르의 슬픔> (1774)

요한 볼프강 폰 괴테
서간체 형식

- 감수성 풍부한 청년 '베르테르'의 사랑 고백
- 시대를 초월해 깊은 공감을 불러일으킨 베스트셀러
- 'Sturm und Drang' 사조의 대표작
- 감정을 자유롭게 풍부하게 표현
- 강렬하고 짙은 감정묘사가 주를 이뤄, 프로젝트 주제에 적합하다고 판단하여 선정함



Natural Language Analysis
with Python NLTK

concordance, word-tokenizer, Text
,lemmatizer, pos-tagger, freqdist
,stopwords, re, numpy
(텍스트 및 단어 분석)



word2vec, LDA 모델
(토픽 모델링)



HUGGING FACE

sentiment-analysis
,text-classification
(감정 분석)

Pipeline

sentiment-analysis dataset
(감정 분석 구체화 데이터)

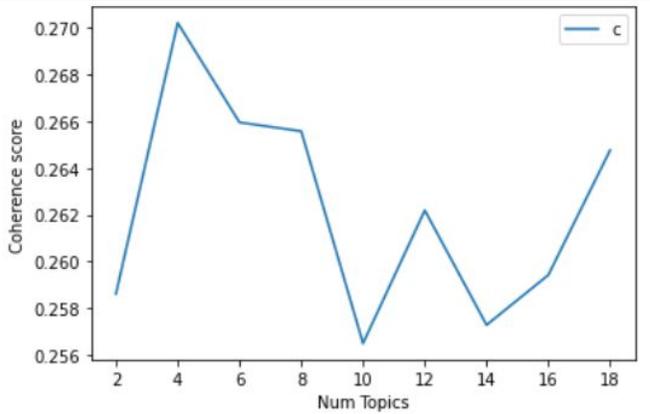


pyplot
(결과 시각화)

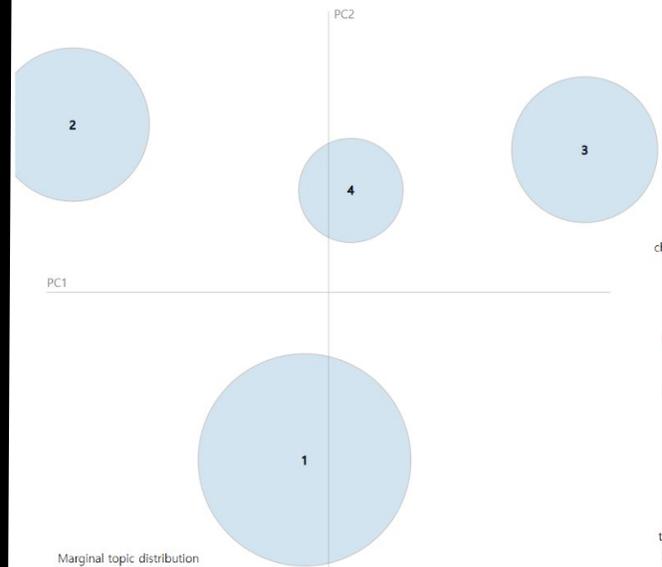
1. 토픽 파악:

dispersion plot과 워드 임베딩을 결합한 토픽 모델링

토픽모델링 결과



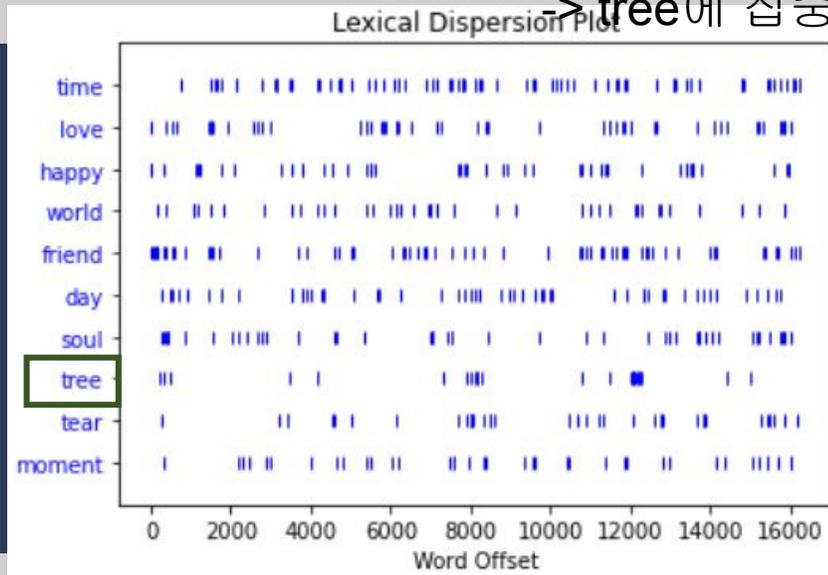
Intertopic Distance Map (via multidimensional scaling)



-2번 토픽-

해당 토픽에서 특징적인 happy, world, moment 모두 특별한 사항 없이 고른 분포

-> tree에 집중하여 토픽 추측



```
[ ] 1 model.wv.most_similar('tree')
[('strength', 0.9995765686035156),
 ('informed', 0.9995686411857605),
 ('cold', 0.9995648860931396),
 ('country', 0.9995511770248413),
 ('after', 0.9995399117469788),
 ('dark', 0.9995396137237549),
 ('opinion', 0.9995395541191101),
 ('letter', 0.9995275735855103),
 ('written', 0.9995246529579163),
 ('storm', 0.9995235204696655)]
```

“사랑의 감정이 죽음으로 이어지는
+ 언어가 가지는 **과정**” 상징적 의미 포착

▶ 3~4개 토픽에서 공통적으로 도출된 단어
: charlotte, man, know, say, feel, little



1. 토픽 파악:

dispersion plot과 워드 임베딩을 결합한 토픽 모델링

토픽 추측 결과

- 1) 'Charlotte'의 집
- 2) 사랑의 감정이 죽음으로 이어지는 과정
- 3) 죽음의 정서
- 4) 죽음을 받아들이는 과정

▶ 별도 조건을 상정하지 않았음에도 원작의 흐름과 유사한 순서로 토픽이 도출되었다는 특이점 발견

젊은 베르테르의 '사랑'이 아닌
'슬픔'인 이유 뒷받침



2. 감정 분석: 2) text-classification

베르테르가 샤를로테를 언급하는 문장에서의 감정 알아보기 pipeline의 text-classification을 이용한 감정 경향성 분석 및 산점도 시각화



긍정/부정을 오가는 결과

- 베르테르의 감정 기복이 드러남

연속해서 **positive**가 나오는 구간

- 샤를로테가 베르테르의 상상 속에서만 등장함.
- 실제 샤를로테와 만나는 장면은 부정값이 섞여 있음

후반으로 갈수록 부정의 비율이 증가

- 자살을 결심한 이후(110-120째 문장) **negative**가 훨씬 자주 등장

▶ Charlotte가 등장하는 문장을 대상으로 실행한 pipeline 산점도



2. 감정 분석:

3) arpanghoshal/EmoRoBERTa 모델

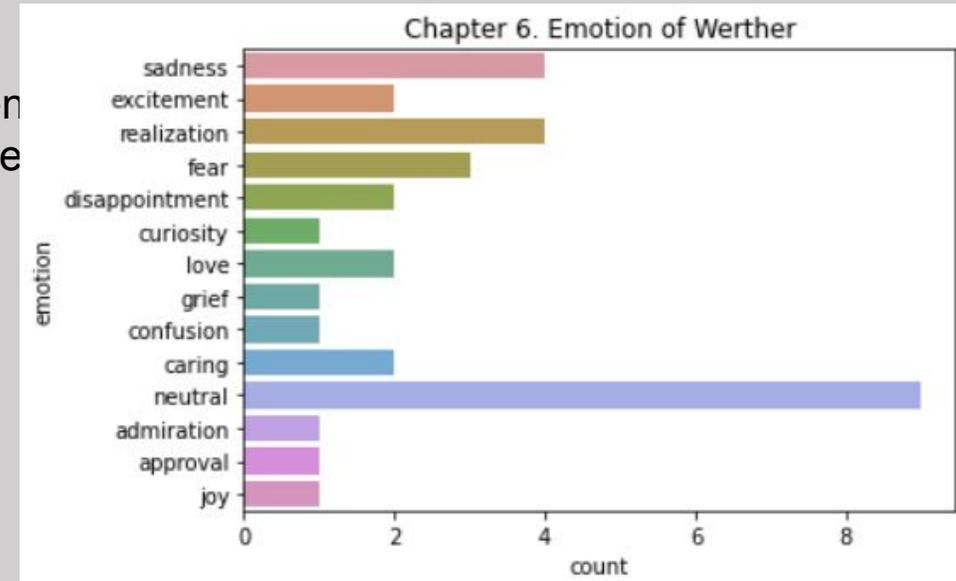
1) 모델 설명

admiration, amusement, anger, annoyance, approval, caring, confusion, curiosity, desire, disappointment, disapproval, disgust, embarrassment, excitement, gratitude, grief, joy, love, nervousness, optimism, pride, realization, relief, remorse, sadness, surprise + neutral

output = [{'label': 'gratitude', 'score': 0.9964383244514465}]

-하나의 문장에 하나의 감정이 대응하는 형식

-가로축: 문장 개수 / 세로축: 특정 문장에 해당하는 감정



2) 분석 방법

- 7개 파트로 분류
- 느낌표가 있는 문장 반환
- Neutral 배제

```
[ ] 1 model.wv.most_similar('로테')
[('더', 0.9234071373939514),
('수', 0.9231418967247009),
('그런', 0.9227683544158936),
('한', 0.9219236969947815),
('이', 0.9216775894165039),
('속에', 0.9213908910751343),
('!', 0.9212024211883545),
('것을', 0.9210866689682007),
('나를', 0.9193261861801147),
('그녀는', 0.9193015098571777)]

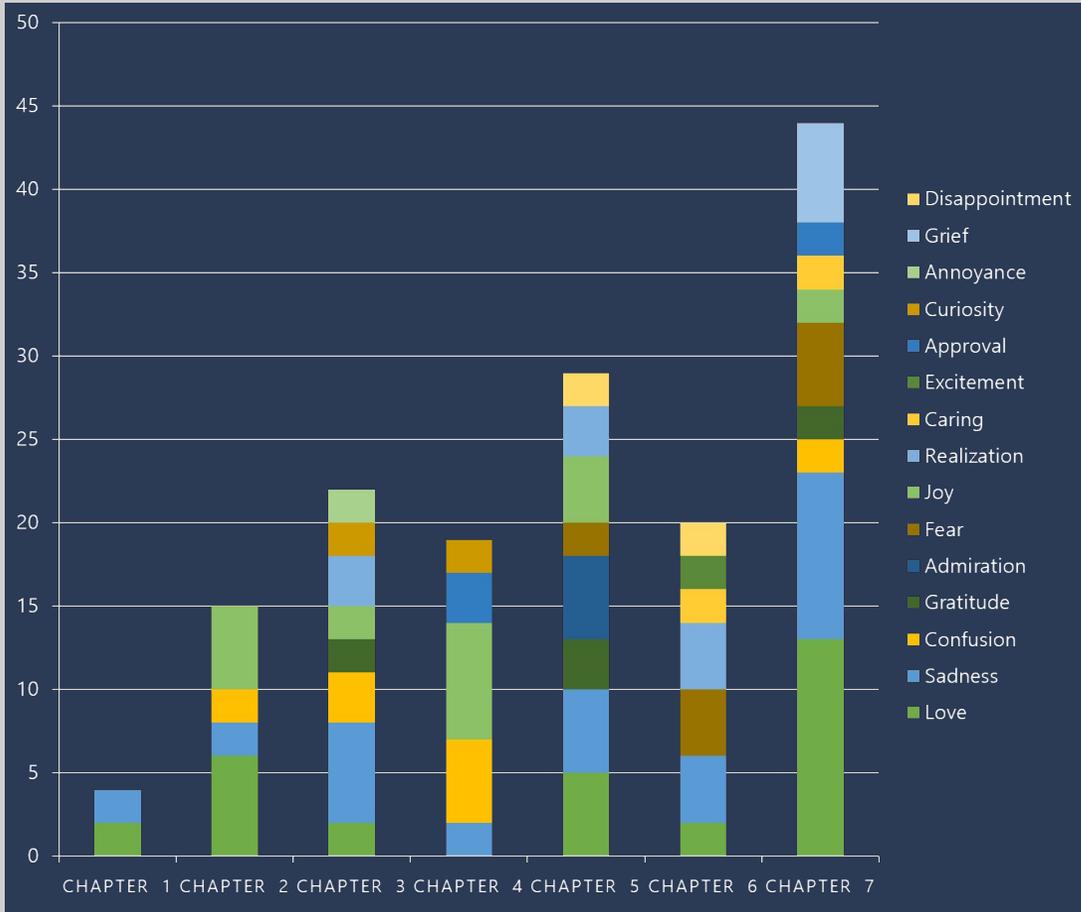
[32] 1 model.wv.most_similar('나')
[('내가', 0.9669130444526672),
(' ', 0.9666948318481445),
('?', 0.9663172960281372),
('!', 0.966251790523529),
('이', 0.9657994508743286),
('그런', 0.9652705192565918),
('그', 0.9635723829269409),
('한', 0.9633512496948242),
('내', 0.962829053401947),
('나는', 0.962364673614502)]

[ ] 1 model.wv.most_similar('알베르트')
[('속으로', 0.5245416164398193),
('지금', 0.517450749874115),
('아직도', 0.5162702798843384),
('서로', 0.5123227834701538),
('정말', 0.506926417350769),
('편지를', 0.5049282312393188),
('!', 0.5019828677177429),
('자신을', 0.4997241193016571),
('그것을', 0.4962400197982788),
('마음을', 0.4959298372268677)]
```

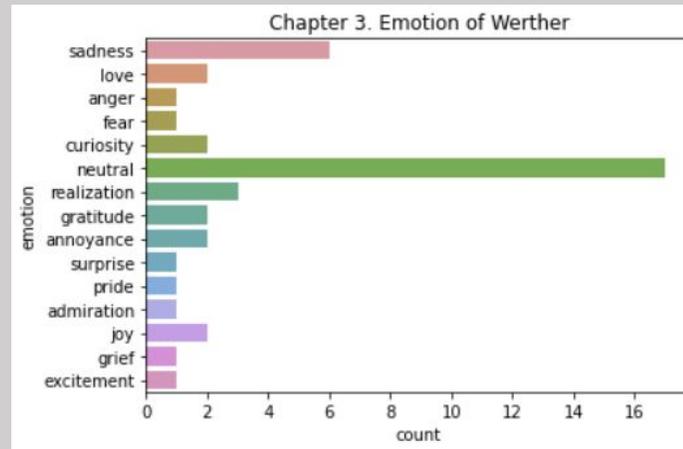


2. 감정 분석:

3) arpanghoshal/EmoRoBERTa 모델



▶ 전 파트 감정 분석 결과



```
[ ] 1 model.wv.most_similar('albert')

[('whom', 0.9993307590484619),
 ('too', 0.9993106126785278),
 ('however', 0.9993070363998413),
 ('past', 0.9992875456809998),
 ('told', 0.9992780685424805),
 ('knew', 0.9992709755897522),
 ('neither', 0.9992702007293701),
 ('quite', 0.9992643594741821),
 ('sweet', 0.9992563724517822),
 ('whether', 0.9992471933364868)]
```

```
1 my_text['content'][1:37].apply(emotion)
```

index	label	score
1	joy	0.9791821241878784
2	neutral	0.9446042180061...
3	confusion	0.99236261844...
4	love	0.92857426404953
5	neutral	0.8109561204910...
6	confusion	0.78487926721...
7	joy	0.9537724256515503
8	love	0.4430732727050781
9	sadness	0.9025776386260...
10	gratitude	0.98176771402...
11	realization	0.459268838...
12	admiration	0.8259884254...
13	joy	0.7770666480064392
14	sadness	0.5461738705635...
15	neutral	0.9875389337539...
16	caring	0.58042675256729...
17	neutral	0.5557172298431...
18	neutral	0.7978556752204...
19	neutral	0.6537305116653...
20	caring	0.35431987047195...
21	neutral	0.4929780960083...
22	love	0.8513208031654358
23	love	0.5174857106398536
24	love	0.9884117245674133
25	love	0.988757312297821
26	fear	0.45500630140304565
27	pride	0.8886181616783142
28	neutral	0.4194501638412...
29	surprise	0.431070059537...
30	joy	0.91835795545578
31	fear	0.4563583731651306
32	love	0.627126796...

- 1) 'Wilhelm: her betrothed is here, -a fine, worthy fellow, whom one cannot help liking' => SADNESS
- 2) 'It would have broken my heart!~I must love him for the respect with which he treats her' => LOVE, ANGER



3. 결론과 한계점

결론과 의의

- ▶ 언어 분석을 통한 감정 분석, 감정 분석을 통한 언어 분석
- ▶ 샤를로테를 향한 사랑과 열정의 증폭이 아이러니하게도 베르테르를 죽음으로 이끈다.
- ▶ 토픽 모델링의 토픽을 결정하는 과정에서 **dispersion plot**, 워드 임베딩의 결과를 종합함으로써 단순 토픽 모델링의 결과만 이용하는 것보다 더 체계적이고 구체적으로 접근하는 방법임을 확인했다.
- ▶ 초반 샤를로테의 언급에서는 **love, joy** 등 순수한 기쁨의 감정이 주로 발견되지만, 후반에는 **love** 와 함께 **sadness, grief, fear** 등 부정적인 감정이 주를 이룬다. 즉, 동일 대상에 대한 호명이라도 그 목적은 다양하다.
- ▶ 파이썬의 자연언어처리로 시행한 언어학적 분석을 인간이 주체가 되어 시행한 문학적 분석과 연결했다는 점과, 문장의 명시적 표현 너머 숨은 감정을 발굴했다는 점에서 학제적인 언어학적 연구로의 의의를 갖는다.

연구의 한계

- ▶ topic modeling과 dispersion plot: 유사한 단어 등장, 고른 분포 -> 확실한 결과 파악 한계
=> 람다값 조정, 명사/형용사 추출, 유사도 분석 시각화, 더 많은 양의 데이터 도입으로 해결 가능
- ▶ 빈도-감정분석: 실제 단어가 가지는 중요성 반영 불가, 감정에 관해서 명백한 결과 X
타 감정분석: 문장 단위 분석, 문장 선정 과정에서의 엄밀성 요구
=> 맥락을 파악한 감정 분석 필요

감사합니다