

텍스트 마이닝을 이용한 소비자 분석 프로젝트

휴먼지능정보공학과 류지혜 이지민

연구 주제

✓ 전자제품의 리뷰로 보는 소비자 심리 연구

▪ 에어팟 1세대와 2세대의 리뷰를 비교하여 아래와 같은 연구를 진행

1. 에어팟 2세대의 변화된 기능을 소비자가 실제로 체감하는 지 리뷰에 대한 텍스트 분석으로 확인
2. 리뷰의 성능 변화 체감 여부에 따라 소비자 만족감에 반영된 심리적 요인을 분석

➡ **비싼 전자제품을 소유하고 있음에 만족 vs 전자제품의 발전된 성능에 만족**

주제 선정 동기

- 전자제품의 출시 이후 1~2년 주기로 기능과 디자인을 리뉴얼, 신규 버전을 시장에 출시하는데, 실제로 기업에서 전자제품의 성능 및 디자인을 개선 한 만큼 소비자들도 그것을 체감하는 지에 대한 호기심이 생김
- 제품에 대한 소비자 평가를 확인할 수 있는 리뷰를 활용하여 연구 진행
- 예시) 애플의 에어팟 2세대: 디자인 차이 X 음성 인식 기능 추가, 배터리 수명 증가 등 성능을 개선하여 출시

SoC	Apple W1
규격	Bluetooth 4.2
센서	가속도 센서, 광학 센서 ^[2]
배터리	케이스: 398mAh (3.81V, 1.52Wh) ^[3] 유닛: 25mAh (93mWh) ^[4]
마이크 개수	각 유닛당 2개 ^[5]
크기	케이스: 44.3 x 21.3 x 53.5 mm, 38 g 유닛: 16.5 x 18.0 x 40.5 mm, 4 g
색상	화이트 ^[6]

에어팟 1세대 사양

SoC	Apple H1
규격	Bluetooth 5
센서	가속도 센서, ^[3] 광학 센서 ^[4]
배터리	케이스: 398mAh (3.81V, 1.52Wh) ^[5] 유닛: 25mAh (93mWh) ^[6]
마이크 개수	각 유닛당 2개 ^[7]
크기	케이스: 44.3 x 21.3 x 53.5 mm, 40 g ^[8] 유닛: 16.5 x 18.0 x 40.5 mm, 4 g
색상	화이트 ^[9]

에어팟 2세대 사양

선행 연구

소비자 텍스트 분석 알고리즘: 텍스트 분석 이론(빈도분석, tf-idf, 감성분석, clustering 등)에 대한 연구와 프로젝트는 계속 진행 중이다.

기존 감성분석 알고리즘은 데이터 수집(빅데이터를 모을 수 있는 웹 크롤링 사용 多) 후, 데이터 토큰화, 불용어제거 등과같은 데이터 전처리 방법을 진행한다. 정제된 데이터를 이용하여 감성분석 알고리즘을 돌려 긍정, 부정도를 `y_train`에 저장한다. 긍정,부정 분류를 위한 LSTM, softmax, categorical_crossentropy 등을 사용한 모델을 구축한다. 이후 정제된 데이터를 훈련(검증 10% 포함), 테스트 데이터로 나누어 지도학습을 진행하여 정확도를 높인다.

소비자선택이론: 소비자의 선택과정에서의 다양한 요인들에 대한 연구가 진행중이다.(보유효과, 매몰비용, 심적회계원리 등)

프로젝트 현재상황

데이터 수집방법, 데이터 분석방법 등에 대한 프로젝트 기획, 기본 웹 크롤링 코드 구현

연구 방법

- 데이터 수집 : 웹 크롤링 (python3 의 BeautifulSoup4 , Selenium 라이브러리 사용)

에플 에어팟 1세대 (MMEF2KH/A) 최저가사라가기

판매처 172 | 제품정보 | 쇼핑물리뷰 13,911 | 쇼핑컨텐츠 | 추천상품

랭킹순 최신순 | 포토&동영상리뷰만 보기

전체 (13,911) | 5점 (11,707) | 4점 (1,606) | 3점 (394) | 2점 (58) | 1점 (106)

주제전제 | 가격 | 만족도 | **음질** | 디자인 | 품질 | 성능 | 기능 | 사용성 | 착용감 | 조작성 | 제품상태 | 송전

★★★★★ 5 열꽃대 · kyhe**** · 19.03.28

배송은 4시 이전 주문 당일 발송으로 다음날 바로 받았구요. 바로 정품 등록했는데 이상 없이 잘 됐어요!~...
배송은 4시 이전 주문 당일 발송으로 다음날 바로 받았구요. 바로 정품 등록했는데 이상 없이 잘 됐어요!~연결도 잘되고, 음악 잠깐 들었는데 음
질도 좋고 정말 편해요^^미리 구입해둔 케이스도 장착해서 바로 사용했어요~가격도 L.POINT랑 청구할인 받아서 직구보다 저렴하게 잘 구입했
습니다. 좋은 상품 감사합니다:)

리뷰펼치기

★★★★★ 5 하이마트쇼핑몰 · b*g***** · 19.05.24

최고의 만족도 에어팟 정말 좋습니다.

아이폰과 아이패드들 둘 다 사용하고 있는데 아이폰에서 음악을 듣거나 영상을 보다 아이패드로 바뀌어서 들을때 불편해서 블루투스 이어폰을 검
색하다 주변의 구매자들 모두 하나같이 최고더라고 이야기 한 애플의 에어팟을 구매하였습니다.

할인쿠폰과 카드청구할인 혜택으로 저렴하게 구입할 수 있었고 배송도 공휴일이 포함되었음에도 2일만에 받을 수 있어 좋았습니다.

리뷰펼치기

1. 네이버 쇼핑에 '에어팟 1세대' 검색

2. 리뷰 주제 (음질 , 품질 , 성능 , 기능) 선택하여 구체적인 평이 반영된 리뷰만을 선택한다.

3. 위 과정을 에어팟 2세대에도 진행

연구 방법

■ 데이터 전처리 :

```
stopwords = ['의', '가', '이', '은', '들', '는', '좀', '잘', '강', '과', '도', '를', '으로', '자', '에 '
```

```
import konlpy
from konlpy.tag import Okt
okt = Okt()
X_train = []
for sentence in train_data['title']:
    temp_X = []
    temp_X = okt.morphs(sentence, stem=True) # 토큰화
    temp_X = [word for word in temp_X if not word in stopwords] # 불용어 제거
    X_train.append(temp_X)

X_test = []
for sentence in test_data['title']:
    temp_X = []
    temp_X = okt.morphs(sentence, stem=True) # 토큰화
    temp_X = [word for word in temp_X if not word in stopwords] # 불용어 제거
    X_test.append(temp_X)
```

4. 크롤링으로 수집된 리뷰를 불용어 토큰화 한 후, 불용어 제거를 한다.

연구 방법

- 데이터 수집 : 웹 크롤링 (python3 의 BeautifulSoup4 , Selenium 라이브러리 사용)

1	종류	카테고리	리뷰	별점
2	에어팟 1세대	품질		
3	에어팟 1세대	품질		
4	에어팟 1세대	품질		
5	에어팟 1세대	품질		
6	에어팟 1세대	품질		
7	에어팟 1세대	성능		
8	에어팟 1세대	성능		
9	에어팟 1세대	성능		
10	에어팟 1세대	성능		
11	에어팟 2세대	품질		
12	에어팟 2세대	품질		
13	에어팟 2세대	품질		
14	에어팟 2세대	성능		
15	에어팟 2세대	성능		
16	에어팟 2세대	성능		

> .csv 파일 예시

5. 웹 크롤링으로 약 7만개 리뷰 중 제품 사양과 정제된 리뷰를 모두 수집

6. 하나의 리뷰에 대하여 [제품 종류] , [카테고리] , [리뷰] , [별점] 순의 데이터를 사진과 같이 .csv 형식으로 저장

연구 방법

- 데이터 분석 : 지도학습 기반 리뷰 긍정/부정 예측 모델 (python3 의 Konlpy, Keras 라이브러리 사용)

	A	B	C	D	E	F
1	종류	카테고리	리뷰	긍정도	부정도	별점
2	에어팟 1세대	품질		0.8	0.2	4
3	에어팟 1세대	품질		0.2	0.8	1
4	에어팟 1세대	품질				
5	에어팟 1세대	품질				
6	0	에어팟 1세대 품질에 대한 80% 만족 에어팟 1세대 품질에 대한 80% 불만족				
7	0					
8	0					
9	에어팟 1세대	성능				
10	에어팟 1세대	성능				
11	에어팟 2세대	품질				
12	에어팟 2세대	품질				
13	에어팟 2세대	품질				
14	에어팟 2세대	성능				
15	에어팟 2세대	성능				
16	에어팟 2세대	성능				

> 감성분석 예시

- ✓ 사전 기반 감성분석 알고리즘을 통해 긍정도, 부정도를 수치화 (이 때, 별점과 긍정도가 상이한 리뷰는 제외한다.)
- ✓ 감성분석으로 얻은 긍정도, 부정도는 지도학습 기반 리뷰 긍정/부정 예측 모델의 target(label)이 된다.
- ✓ 예측 모델 결과로 얻은 긍정/부정도를 1세대와 2세대 별로 비교하고, 1세대와 2세대의 별점, 판매량과 비교한다.

예상 결과

■ Review 분석 결론

- 성능 개선이 소비자 만족도에 영향을 준다. Or 성능 개선과 소비자 만족도는 관련이 적다.
- 에어팟 1세대와 2세대의 평균 별점

✓ 소비자가 전자제품에 만족도를 느낀 요인이 심적 회계원리나 보유 효과에 의한 것인 지, 개선된 성능 자체에 만족한 것인지 분석

- 에어팟은 가격대가 높기 때문에 소비자가 심적 회계원리에 의해 긍정적인 리뷰를 남겼을 가능성이 있음
- Review는 제품을 이미 구매한 사람이 쓴 것이기 때문에 보유효과에 의한 긍정적인 리뷰를 남겼을 가능성이 있음
- 예시) 1세대에 비해 2세대가 성능 개선과 관련된 긍정 리뷰는 적으나, 판매량과 전체 평점이 증가했다면 소비자는 좋아진 성능 보다는, 비싼 기기 값에 의한 심적회계원리와 보유효과 등에 의해 만족감을 느껴 긍정 리뷰를 작성했다.