



온라인 리뷰 특화 모델 : ReBERT, ReELECTRA



25.11.03-11.14



TEAM3



온라인 리뷰 특화 모델

목차

프로젝트 소개

데이터셋

모델 설명

성능 평가 및 결과

감성 분석 활용 사례

소감

TEAM 3



목 차

Chapter.1

프로젝트 소개

Chapter.2

데이터셋

Chapter.3

모델 설명

Chapter.4

성능 평가 및 결과

Chapter.5

감성 분석 활용 사례

Chapter.6

소감



온라인 리뷰 특화 모델

목차

프로젝트 소개

데이터셋

모델 설명

성능 평가 및 결과

감성 분석 활용 사례

소감



프로젝트 소개



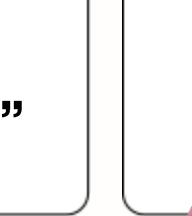
“전자상거래 시장 규모 확대”



“기업의 목적은
고객을 창조하는 것”



“타 기업과의 차별점을 위해
온라인 고객 리뷰(OCR) 분석은 필수”



“리뷰는 잠재 고객의
구매 의사 결정에 영향”

----- How is OCR used? -----

온라인 고객 리뷰 분석의 필요성 증가

TEAM 3



온라인 리뷰 특화 모델

목차

프로젝트 소개

데이터셋

모델 설명

성능 평가 및 결과

감성 분석 활용 사례

소감

TEAM 3



프로젝트 소개



Goals

온라인 리뷰 특화 한국어 NLP 모델 구현!

NLP 모델

- BERT, ELECTRA
 - 문어체, 구어체 등 다양하게 수집한 데이터셋 300만개
- 사전학습

DAPT

- 특정 도메인의 데이터셋 추가 학습
- 해당 도메인에 대한 이해도와 성능 향상

감성 분석

- 텍스트의 긍부정 분류
- 언어 속 감정 신호를 정량화하여 데이터 기반 의사 결정 지원



온라인 리뷰 특화 모델

목차

프로젝트 소개

데이터셋

모델 설명

성능 평가 및 결과

감성 분석 활용 사례

소감

TEAM 3



데이터셋

학습 데이터

Pre-training Data

- 뉴스 리뷰(GitHub), 일상 구어체(AIHUB), 뉴스 기사(AIHUB), 한국 문화 기사, 웹사이트 설명, 게임 리뷰, 영화 리뷰
- 총 약 300만 건의 텍스트 데이터

DAPT

- 특정 도메인의 데이터를 활용하여 사전 학습된 언어 모델의 성능을 개선
- 쇼핑몰 리뷰 22 만 건(AIHUB)

Fine-tuning Data

- 온라인 쇼핑 리뷰에서의 모델 성능 향상을 위한 네이버 쇼핑 리뷰 20만 건

타겟 데이터

- 크롤링을 이용한 수집
- 온라인 쇼핑 플랫폼
(무신사)에서 최신 리뷰 데이터 수집
- 총 12만 건

데이터 전처리

- 모든 데이터는
'5 가지 데이터 정제 규칙'에 따라 전처리



온라인 리뷰 특화 모델



목차



프로젝트 소개



데이터셋



모델 설명



성능 평가 및 결과



감성 분석 활용 사례



소감

TEAM 3

BERT

MLM

오늘 날씨가 좋아서 산책을 나갔다.

Tokenizer

[CLS] 오늘 날 ##씨 가 좋 아 서 산 책 을 나 갔 다 [SEP]

'좋','산','갔' token 변형

[CLS] 오늘 날 ##씨 가 [MASK] 아 서 문 책 을 나 갔 다 [SEP]

- Transformer 기반 모델
- 기존의 RNN, LSTM 계열 모델과 달리 양방향 문맥을 모두 고려할 수 있음
- Pre-training 단계에서 MLM을 통해 언어 학습
- Token 중 15%를 마스크 처리
- 80%를 [MASK], 10%를 다른 단어, 나머지 10%를 유지



온라인 리뷰 특화 모델



목차



프로젝트 소개



데이터셋



모델 설명



성능 평가 및 결과



감성 분석 활용 사례



소감

TEAM 3

ELECTRA

RTD

오늘 날씨가 좋아서 산책을 나갔다.

Tokenizer

[CLS] 오늘 날 ##씨 가 좋아서 산책을 나갔다 [SEP]

'좋','산','갔' token 변형

[CLS] 오늘 날 ##씨 가 [MASK] 아서 문책을 나갔다 [SEP]

변형된 토큰 예측
(Generator)

[CLS] 오늘 날 ##씨 가 밥 아서 문책을 나갔다 [SEP]

토큰의 진위를 예측
(Discriminator)

[CLS] 오늘 날 ##씨 가 밥 아서 문책을 나갔다 [SEP]

↑ fake ↑ fake

- 기존 BERT의 MLM을 대체하기 위해 개발된 모델
- Pre-training 단계에서 RTD를 통해 언어 학습
- RTD는 Generator(생성기)와 Discriminator(판별기)로 구분
- BERT보다 더욱 효율적인 학습이 가능



온라인 리뷰 특화 모델

목차

프로젝트 소개

데이터셋

모델 설명

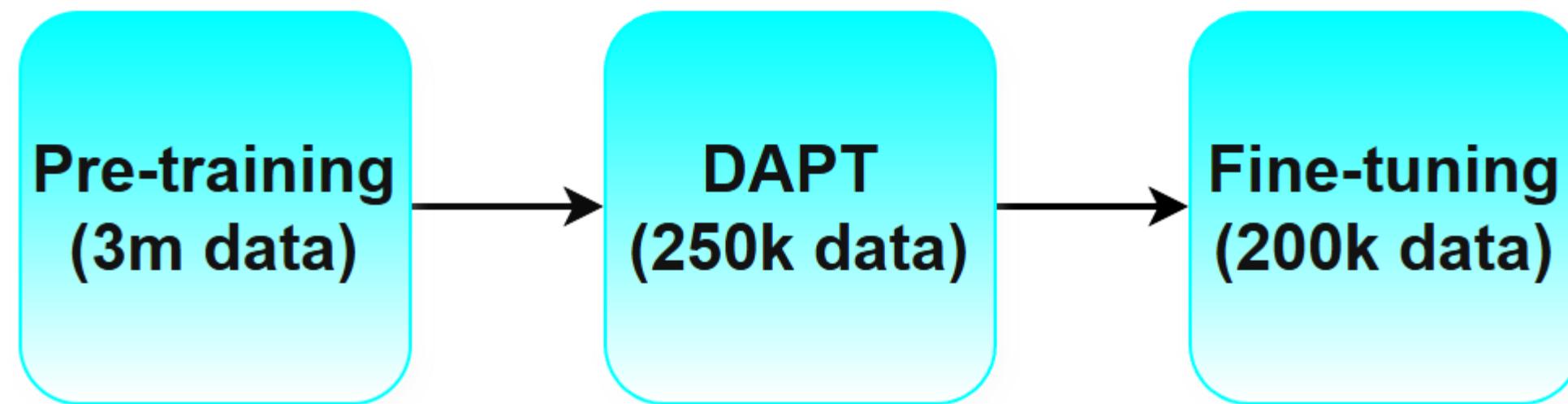
성능 평가 및 결과

감성 분석 활용 사례

소감

TEAM 3

ReBERT, ReELECTRA



200만개의 구어체
100만개의 문어체

쇼핑몰 리뷰 데이터

네이버 쇼핑 리뷰
댓글 데이터



온라인 리뷰 특화 모델

목차

프로젝트 소개

데이터셋

모델 설명

성능 평가 및 결과

감성 분석 활용 사례

소감

TEAM 3



성능 평가 및 결과

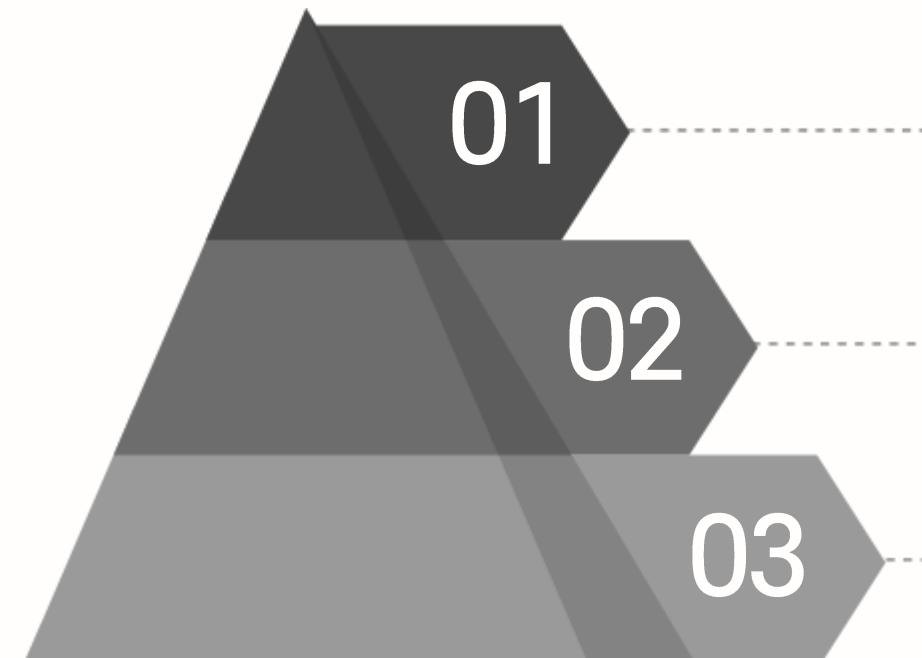


모델 성능 평가

Model	KcBERT	ReBERT	KcELECTRA	ReELECTRA
Size	Base	Small (tuning)	small	Small
Accuracy	93.41 %	89.53 %	91.70 %	88.98 %
F1-Score	94.62 %	91.53 %	93.21 %	90.90 %



기존 모델과의 차별점



- 경량성 (Lightweight)

- 효율성 (Efficiency)

- 도메인 최적화 (Domain Adaptation)



온라인 리뷰 특화 모델

목차

프로젝트 소개

데이터셋

모델 설명

성능 평가 및 결과

감성 분석 활용 사례

소감

TEAM 3



성능 평가 및 결과



한계점

- 1 제한된 자원 환경과 시간적 제약으로 인한 학습 한계
- 2 학습 데이터의 규모 확대 필요성
- 3 하이퍼 파라미터 최적화 과정의 제한 (시간적 제약)



개선 방향

- 충분한 컴퓨터 자원 확보
- 대규모 리뷰 데이터셋 추가 확보
- Learning Rate, batch size, hidden layer size 등 하이퍼 파라미터 튜닝 필요



온라인 리뷰 특화 모델



프로젝트 소개



데이터셋



모델 설명



성능 평가 및 결과



감성 분석 활용 사례



소감

TEAM 3



감성 분석 활용 사례



인사이트 도출

소비자 구매 의사결정 요인

→ 심미성보다 실용성 중심의 소비 경향

제품 개선 및 마케팅 포인트

→ 소비자가 중요하게 여기는 키워드를 중심으로 핵심 품질 관리 및 세분화 전략

카테고리 차별화 전략

→ 후드티의 경우 핏과 사이즈의 세분화나 소재 설명 강화로 차별화





온라인 리뷰 특화 모델

목차

프로젝트 소개

데이터셋

모델 설명

성능 평가 및 결과

감성 분석 활용 사례

소감

TEAM 3

곽동원



처음에는 새소하고 어려운 개념들이 많았지만, 팀원들과 함께 하나하나 토론하고 정리해 나가면서 좋은 공부가 되었고, 토론하며 정리하는 자체가 큰 배움이었습니다. 2주라는 짧은 기간 동안 진행된 프로젝트였지만, 각자 맡은 역할에 책임감을 갖고 성실히 임해준 덕분에 좋은 결과를 만들어낼 수 있었고 협업의 즐거움을 다시금 느낄 수 있었습니다.

안호용



2주라는 기간동안 팀원들과 같이 프로젝트를 했는데, 어려운 주제를 팀원들과 같이 토론하면서 풀어나가는 과정이 참 좋았습니다. 이렇게 오랫동안 팀 프로젝트를 해본 적이 없어서 걱정을 많이 했는데, 그럼에도 다들 열심히 해주셔서 이렇게 좋은 결과물이 나올 수 있었네요. 다들 수고하셨습니다!

한국어 사전학습부터 진행한 덕분에 모델의 동작 원리와 구조를 자세하게 알아보고 이해할 수 있어서 좋은 공부가 되었습니다.

모든 팀원이 각자 맡은 역할에 충실하고 밤낮없이 열정적으로 임해주셔서 짧은 기간이었지만 좋은 결과가 나올 수 있었던 것 같습니다.

강지연



수업 중 배웠던 딥러닝 모델 외에 두 가지 모델을 추가로 배우고 작동 원리와 구조에 대해 공부할 수 있어 유익한 시간이었습니다. 프로젝트 주제를 선정하면서부터 과정이 쉽지 않을 것 같아 걱정이 있었는데 좋은 팀원들을 만나 모두 열정을 가지고 임하는 모습에 동기부여가 많이 되었고, 최종적으로 결과물도 잘 나온 것 같습니다.

다들 수고 많으셨습니다~!!

정수아





감사합니다



Q & A