

딥러닝 기반 영화평 감성 분석

[리뷰 해체 분석기]

Team #4

201411273 박재범

201411275 박진호

201411283 이상민

201511244 김민우

1. 작품 소개
2. 사용 SW/HW
3. 최종 산출물의 형태 및 기능
4. **Alternative Solution(Reference) & Project Justification**
5. **Risk Analysis & Risk Reduction Plan**
6. **Success Criteria**

작품명: 리뷰 해체 분석기

설명:

문화체육부와 국립국어원에서 공동 주최하는 국어 정보처리 시스템 경진대회의 2020년도 주제인 '딥러닝 기반 영화평 감성 분석' 주제를 바탕으로, 구글의 AI 언어모델 BERT 기반 영화평 분석 기계학습 모델을 개발한다. 한글로 작성된 영화평을 읽어들이어 해당 영화평의 긍정/부정을 분류하는 것이 목적이며, BERT에 독자적으로 개발한 아키텍처를 융합하여 성능 및 정확도를 향상시킨 새로운 파생 언어모델의 완성을 프로젝트의 최종 목표로 한다.

1. 작품 소개(시나리오)

id	document	label
0 9976970	아 더빙.. 진짜 짜증나네요 목소리	0
1 3819312	졸...포스터보고 초딩영화줄...오버연기조차 가볍지 않구나	1
2 10265843	너무재밌었다그래서보는것을추천한다	0
3 9045019	교도소 이야기구면 ..솔직히 재미는 없다..평점 조정	0
4 6483659	사이몬페그의 허클베리핀 연기가 돋보였던 영화스파이더맨에서 늙어보이기만 했던 커스틴 ...	1

```
print(train_pos[0])

['아/Exclamation',
 '더빙/Noun',
 './Punctuation',
 '진짜/Noun',
 '짜증나다/Adjective',
 '목소리/Noun'],
 '0']
```

review	positive	negative	mind
1 일본 특유의 감성이... 다 우리나라에선 절대 볼 수...	2	0	
2 정말 한번쯤은 봐야 하는 인듯 기대하고봐서 눈물은 생각보...	1	0	1
3 순수하고 맑은 사랑을 본 느낌이었어요	1	0	1
4 두번보면 안되는	0	0	0
5 여자가 너무 불쌍하다 남자는 여자에게 말하면서 좋기라도 ...	0	1	-1
6 이 일본영화 또 왜이렇게 아련한거야	0	0	0
7 일본판 이프온리 길으면서도더 심오하고 가슴아픈 서로가 ...	0		
8 너무 이쁘고 순수하고 슬픈사랑 이야기 내가 저 상황이면 다...	1		
9 역시 일본은 이종적이다 그래서 싫지만 좋다	0		
10 역시 사랑에 노력은 필요한거구나 상대방을 생각하는 배려 ...	0		

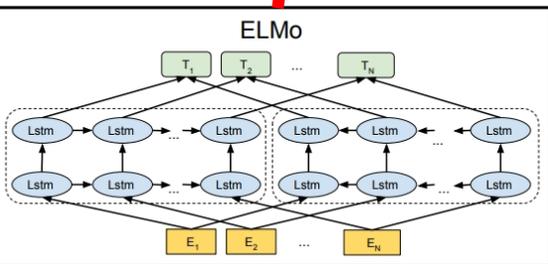
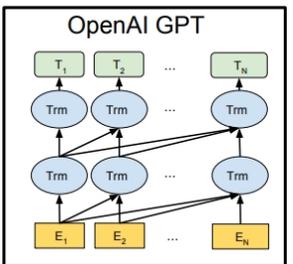
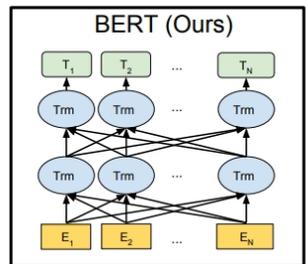
가난
가난뱅이
가난살이
가난살이하다
가난설음
가난에
가난에 쪼들려서
가난하게
가난하고
가난하고 어렵다
가난하고 천한
가난하다
가난하다

XD
가시
가도
가격이 비싸
가까이 사귀어
가까이하다
가다듬어
가다듬어 수습하는
가다듬어 수습하다
가다듬어 정하다
가당히
가뜩하게
가리안섬 바르자다

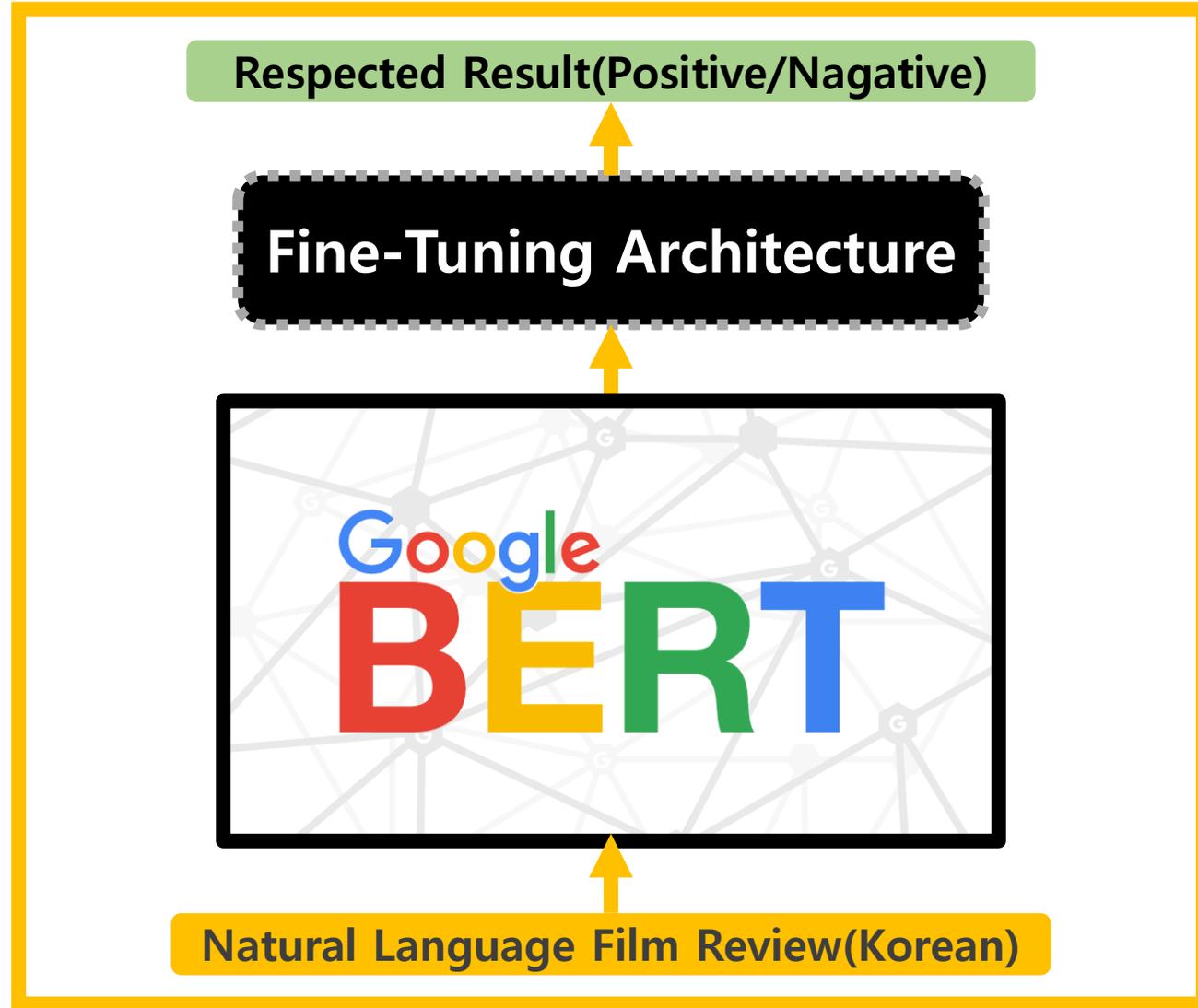
1

3

2



① 자연어로 쓰여진 영화평 데이터에 대해 ② ELMo(Word Embedding)를 통해 ③ 부정/긍정의 도출



▲ Overall Structure

새로 만들 SW:

BERT 모델로부터 생성한 데이터들을 Fine-Tuning하여
성능 및 효율성을 높이는 아키텍처 및 인터페이스

COTS SW:

BERT, PyTorch 또는 TensorFlow

HW: Windows 10 OS 기반 Computer



형태:

BERT 기반 RNN을 이용한 언어모델

기능:

한글로 작성된 영화평을 읽어들이어
해당 영화평의 긍정/부정을 분류

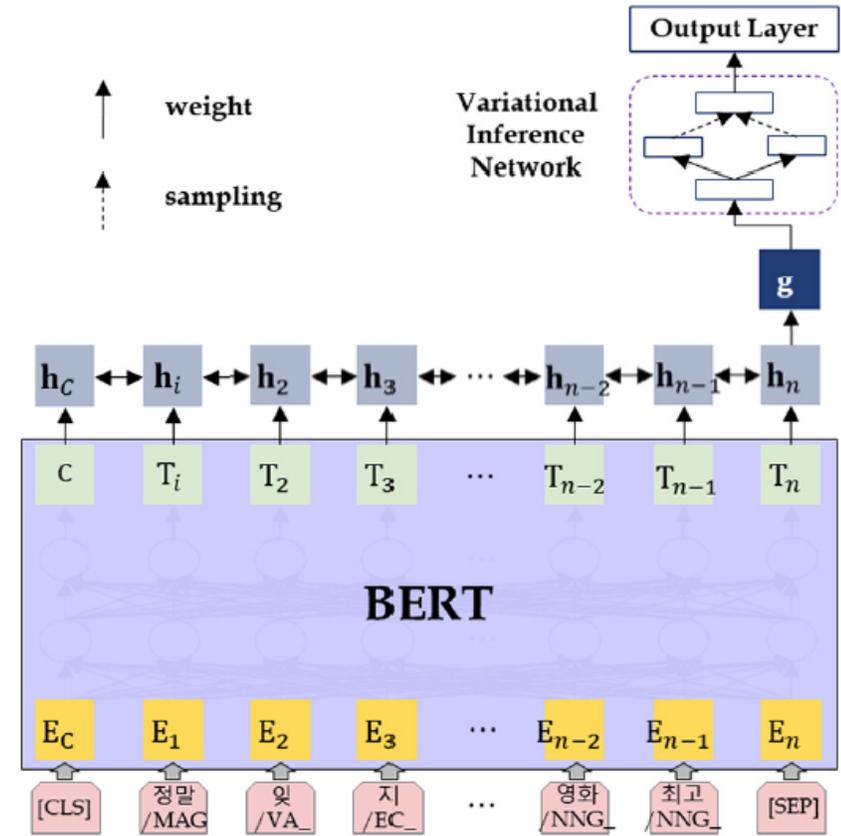


그림 4 BERT 모델 기반 RNN을 이용한 variational inference 감성 분석 모델

Fig. 4 Sentimental analysis model with variational inference using RNN based on BERT

출처: BERT 기반 Variational Inference와 RNN을 이용한 한국어 영화평 감성 분석(<http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE09233106>)

Alternative Solution(Reference):

KoBERT, RoBERTa, XLNet, ALBERT, ELECTRA

Project Justification:

- **Cost:** 오픈소스 SW 활용으로 저렴한 비용
- **Duration:** 약 3~4개월
- **Risk:** 짧은 기간동안 많은 학습량이 필요
- **Effect:** 개발 비용 절약, 팀원 개발 능력의 정량적 분석 가능,
기존 모델들의 장단점 분석을 통해 향상된 성능 및 정확도의
자연어 처리 언어모델 도출

Risk	Probability	Significance	Weight	Reduction Plan
딥러닝에 대한 이해 부족	5	5	25	논문 및 온라인 강의를 통해 열심히 학습
짧은 프로젝트 기한(10월 초)	5	4	20	효율적인 일정 수립
COTS SW의 활용 미숙	3	3	9	방학 중 PyTorch 세미나 참여
언어모델 이해 부족	4	5	20	논문 및 온라인 강의를 통해 열심히 학습
갑작스러운 팀원의 취업	1	1	1	역할 분담과 인수인계를 통한 후속 조치
취업 준비	5	2	10	적절한 비중 안배
근손실	3	5	15	고중량 저반복을 통해 훈련 시간을 단축

성능:

BERT 기반 언어모델의 평균 프로세싱 시간을 5% 이상 단축

정확도:

BERT의 정확도보다 0.5%p 이상 향상

성과:

2020 국어 정보처리 시스템 경진대회 45위 이상 달성