

이건 뭐 사람 아니냐고

BERT 기반의 영화평 감상 분석

201310737 김주언
201611829 최보라
201712845 류한길
201411283 유병찬

Description

“이건뭐사람아니냐고”는 어떤 프로그램인가?



Google 에서 공개한
인공지능 언어모델인 BERT를 기반으로
영화평 감상을 분석해주고
주어진 문장으로만 그 문장이 긍정적인 문장인지
부정적인 문장인지 판단하는 프로그램

BERT 기반의 영화평 감상 분석

SW/COTS SW/HW

Description

SW/COTS SW/HW

최종 산출물 형태 및 기능

새로 만들 SW

영어를 기반으로 만들어진 기존의 BERT를 이용하여 한국어 기반의 모델을 만든다.

Alternative Solution

Project justification

COST SW

HanBERT, ALBERT, Electra, KNU 한국어 감성 사전

Risk Analysis

Risk Reduction Plan

Success Criteria

최종 산출물 형태 및 기능

1. 영화평 문자열 긍/부정 분석 프로그램

- 문장만으로 그 문장이 긍정적인 문장인지 부정적인 문장인지 파악할 수 있다. state-of-the-art 수준의 기계학습 모델을 개발한다.

2. 시연용 웹 프로그램

- 전시 참여자가 영화평을 입력하면, 긍정/부정을 판단해줄 수 있게 구현
- 가장 최적화된 프로그램을 만들기 전에 시도했던 여러 모델의 성능을 시각적으로 비교하고, 각 모델을 설명해주는 페이지 구현

Alternative Solution

- 네이버 영화 리뷰 감정 분석

[https://colab.research.google.com/drive/1tlf0Ugdqg4qT7gcxia3tL7und64Rv1dP#scrollTo=P58qy4--s5_x]

- 가장 유명한 모델 중 하나인 Hugging Face의 Pytorch BERT를 사용
- BERT의 기본 모델인 bert-base-multilingual-cased 를 그대로 이용하기 때 문에 한글case에서 정확도가 낮게 나옴

Project justification

- BERT의 기본 모델인 bert-base-multilingual-cased 를 사용하는 것이 아니라 한국어에 딱 맞는 모델을 설계하여 정확도를 높인다.
- 한국어 감성사전을 Feature 로 추가하여 정확도를 높인다.
- 기존에 있는 다양한 BERT 를 사용하는 앙상블 기법으로 정확도를 높인다.

Risk Analysis

- 학부에서 배울 기회가 없어서 딥러닝과 신경망 모델에 대한 이해도가 낮은 팀원들이 존재
- 모델을 학습시키는데 시간이 오래 걸리고, 또 학습 후에 시행착오를 겪으며 모델을 수정하는데 시간이 또 들어 시간이 부족할 수 있음
- 시간 부족 문제에 따라서 우리가 원하는 만큼의 정확도를 내기 어려움

Risk Reduction Plan

Description

- 학부에서 배울 기회가 없어서 딥러닝과 신경망 모델에 대한 이해도가 낮은 팀원들이 존재

SW/COTS SW/HW

-> 최소 주 1회의 스터디를 진행하여 딥러닝 프로젝트 개발 경험이 있는 팀원들의 도움을 받는다. 매주 1회 교수님과의 면담을 진행하여 개발 방향과 공부 방향이 맞는지 자문을 구하고 조언을 받는다.

최종 산출물 형태 및 기능

Alternative Solution

Project justification

- 모델을 학습시키는데 시간이 오래 걸리고, 또 학습 후에 시행착오를 겪으며 모델을 수정하는데 시간이 또 들어 시간이 부족할 수 있음
- 시간 부족 문제에 따라서 우리가 원하는 만큼의 정확도를 내기 어려움

Risk Analysis

Risk Reduction Plan

-> 시간이 모자랄 것이라는 것을 인지했기 때문에 좀 더 빠르게 공부를 시작하였고 앞으로의 프로젝트 진행 계획을 세웠다. 좀 더 자주 팀원과 만나서 프로젝트 진행을 원활히 할 것이다. 시간이 부족한 만큼 교수님께 더 많은 피드백을 받아서 진행 방향이 어긋나지 않게 한다.

Success Criteria

BERT 기반의 영화평 감상 분석

Success Criteria

Description

SW/COTS SW/HW

최종 산출물 형태 및 기능

Alternative Solution

Project justification

Risk Analysis

Risk Reduction Plan

Success Criteria

한국어에 최적화된 모델 개발을 해서
정확도가 91% 이상이 되기