



Upstage AI Lab

NLP Dialogue Summarization 경진대회
Seminar | 2025. 1. 31(금)

목차

01. 팀 소개

02. 경진대회 수행 절차 및 방법

03. 분석 인사이트 및 결과

04. 회고

01

팀 소개

팀장/팀원 소개
협업 방식

Byte Busters

NLP의 기본 단위인 형태소 파악이 중요하듯이, 더 넓게 포괄하는 "데이터"의 기본 단위인 **Byte**를 정복한다.



팀장

신다혜

LLM

데이터 전처리



팀원

이다연

LLM

데이터 전처리



팀원

조혜인

ML, CV

모델링



팀원

송주은

LLM

모델링 및 LLM 활용



팀원

조성지

CV/LLM

모델링, A/B테스트

경진대회 협업 방식

: Natural Language Processing [대회] Dialogue Summarization

- 🎉 오프라인 미팅으로 친목 다지기 🎉
- 논의 방식: Zoom 미팅. 회의록 기록
- 일정 관리: Github Projects 기능 활용
- 코드 및 데이터셋 관리
 - Github
- 학습 결과 공유
 - WandB 조직 내 개별 프로젝트 생성



02

경진대회 수행 절차 및 방법

목표 수립
수행 내용 / 수행 결과

경진대회 목표 수립

: Natural Language Processing [대회] Dialogue Summarization 2025. 01. 15 ~ 01.27

주제

Dialogue Summarization | 일상 대화 요약

목표

- Final 점수 50점
- 데이터 전처리, LLM 모델 등을 적용하여 점진적인 성과 향상

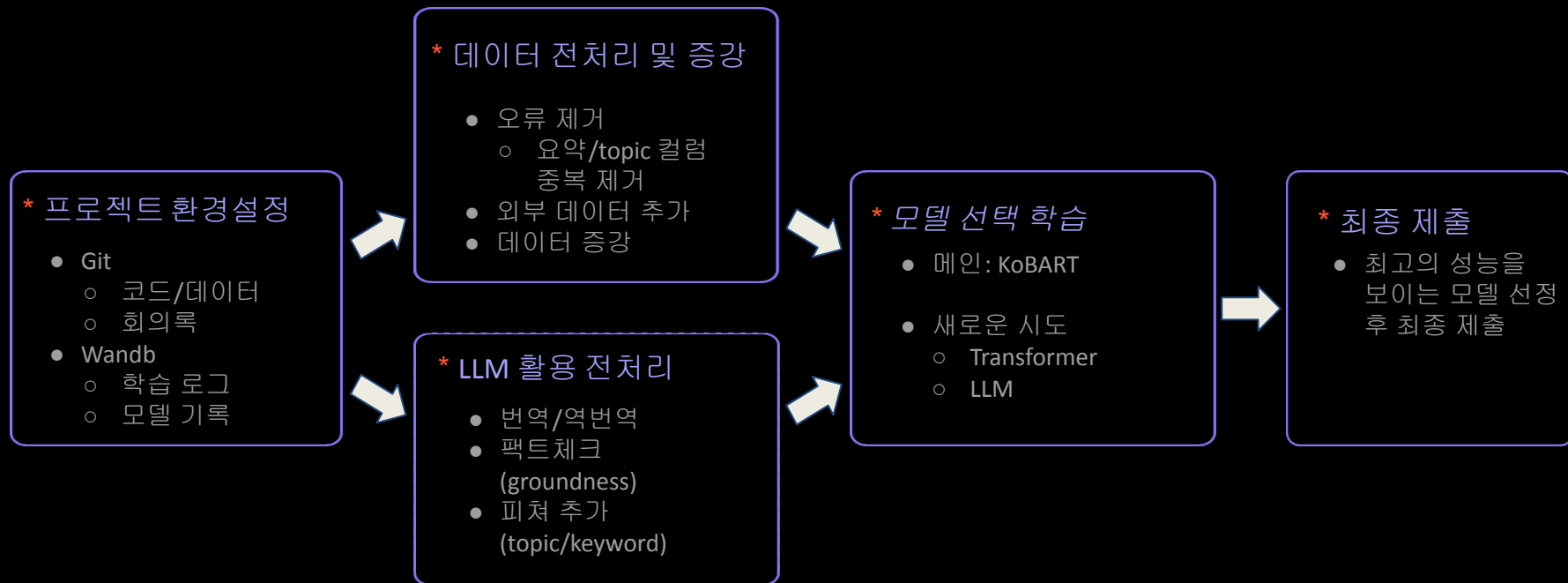
개요

소개 및 배경 설명

- 일상 대화를 효과적으로 요약하는 모델을 개발하는 경진대회
- 제공된 대화문과 요약문 데이터셋을 활용하여, 다양한 자연어 처리 기법을 적용해 요약문을 생성하는 모델 구축
- 비정형 텍스트 데이터 처리 및 자연어 요약 모델링 역량 강화

경진대회 수행 내용

: Natural Language Processing [대회] Dialogue Summarization



03

분석 인사이트 및 결과

문제 및 인사이트 도출
해결 방법 및 결과

경진대회 인사이트 공유

: Natural Language Processing [대회] Dialogue Summarization

KoBART 모델 활용 및 데이터 전처리 진행

- 데이터 전처리
 - 스페셜 토큰 누락
 - 요약 내용 중복 존재
 - 요약 퀄리티가 낮은 학습 데이터 수정
- KoBART 모델 튜닝
 - WandB Sweep으로 튜닝
 - A/B 테스트 진행

새로운 모델 적용 및 전처리 기법 시도

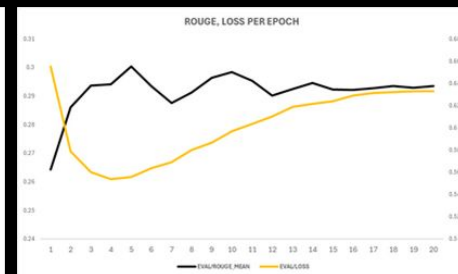
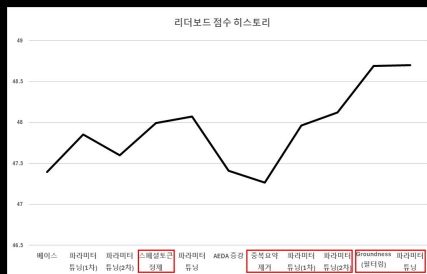
- KoBart의 모델이 어느 일정 수준 점수가 오르지 않는 문제를 해결하기 위해 새로운 모델 시도
 - Transformer
 - T5, Google/Pegasus
 - LLM
 - Upstage/Solar, Google/Gemini
- 효과적인 요약을 위해 토픽/키워드 피쳐 추가
- 기존 데이터셋의 어색한 한국어 문장 구조로 인한 오류 개선 및 영어 특화 모델 활용을 위한 번역 진행
- T5 모델이 스페셜 토큰을 인식 못함

경진대회 인사이트 공유

: Natural Language Processing [대회] Dialogue Summarization

점수 향상 히스토리 개괄

- 데이터 전처리
 - 누락 및 오기입 된 스페셜 토큰 정제
 - 중복된 요약문 중복 제거(16개)
 - Solar Groundness 활용하여 낮은 요약
퀄리티 데이터 삭제(197개)
- A/B 테스트, 하이퍼파라미터 튜닝
 - WandB 사용하여 효율적
하이퍼파라미터 탐색
 - 데이터 변경 후 A/B테스트로 성능 비교
 - Loss 최소화 최적화 -> Rouge 최대화
최적화로 변경



경진대회 인사이트 공유

: 점수 향상 히스토리

Data	스페셜토큰	스페셜토큰 + 중복제거 + 저퀄리티 삭제	스페셜토큰 + 중복제거 + 저퀄리티 삭제 + 증강
LearningRate	1e-5	1.14e-5	2e-5
BatchSize	16	16	50
WeightDecay	0.01	0.017	0.03
GradientAccummulation	1	1	4
BestModel Selection	Loss 최소화	Rouge 최대화	

데이터 클리닝시 가장 높은 성능 향상을 기록했으며 AEDA, 추가 데이터 수집을 통한 증강을 적용했을 때의 성능은 하락,
베스트 모델 저장 기준을 Rouge 최대화로 변경하여 검증 결과의 신뢰도를 높일 수 있었다.

경진대회 인사이트 공유

여러가지 시도 - 모델

오픈소스 모델 (파인튜닝)

상용모델 (LLM API + Prompt Engineering)

 digit82/kobart-summarization

T5

Facebook  Bart Large Cnn

 google

/pegasus-xsum

Solar Mini

Solar Pro



Experimental

Gemini 2.0

경진대회 인사이트 공유

여러가지 시도 - 데이터

데이터 전처리

- Solar-mini로 요약문 팩트체크
 - 300개 제거->점수하락
 - 제거가 아닌 수정했으면 데이터 살렸을 듯
- 외부데이터 확장
 - AI_Hub 한국어 대화 요약 12천개 확장-> 점수하락
- 데이터 증강을 통해 더욱 풍부한 학습 데이터셋을 만들고자함
 - AEDA 방식 적용(문장기호 추가)

번역/역번역

- LLM을 활용한 한-영 번역 및 영-한 역번역
 - 목적:
 - 1) 데이터 특징이 영어를 한국어로 직역한 것 같다는 가정
 - 2) 영어 기반 Transformer 모델 활용
 - 모델 : Solar / Gemini 2.0

토픽/키워드

- Train 데이터와 동일하게 Test 데이터셋에 'topic' 컬럼 추가
 - 모델: Solar-mini
 - i. Train과 Dev의 'topic' 컬럼을 참조하도록 설정
 - ii. Test 데이터셋에도 동일하게 'topic'을 추출하도록 설정
- 요약 가이드로 활용할 키워드 추출
 - 모델: Gemini 2.0(flash/thinking)
 - i. 학습 데이터의 대화문과 요약문에서 공통 키워드 추출
 - ii. Gemini chat-session에 입력하여 test 데이터에서도 키워드 추출

04

회고

우리 팀의 목표 달성도
느낀점 및 향후 계획

경진대회 회고

: Natural Language Processing [대회] Dialogue Summarization

Point 1

우리 팀의 처음 목표에서 어디까지 도달했는가

* 1등 : 50% 의 성공, 나머지 50% 미래를 기약

Point 2

우리 팀이 잘했던 점

* 많은 일을 각자 잘 수행

Point 3

협업하면서 아쉬웠던 점

* 조금 더 많은 모델을 사용해보지 못해서 아쉬움

* 향후 계획 : 허깅 페이스를 통해 많은 모델을 로드해와서 다시 한번 요약문을 생성하는 방식으로 테스트를 진행

경진대회 진행 소감

: Natural Language Processing [대회] Dialogue Summarization



* 신다혜 다양한 모델과 LLM을 사용해볼 수 있어서 좋았습니다.

* 이다연 다음에는 허깅페이스에서 모델 받고 파인튜닝하는 걸 최수선 목표로 할 것이다.

* 조혜인 LLM 사용방법을 깨닫고, 앞으로 LLM을 다룰지에 대한 힌트를 얻을 수 있던 대화였던 것 같습니다

* 송주은 다양한 NLP 모델과 LLM을 활용해보며 익숙해질 수 있는 좋은 경험이었습니다

* 조성지 모델 실험체계와 파라미터 튜닝, A/B 테스트를 진행하면서 점진적으로 모델 성능의 향상을 경험한 것이 의미있었다.

Life-Changing Education

감사합니다.
