

CUAI 동계 컨퍼런스 NLP팀 중간 발표

2023.01.17

발표자 : 김지민

팀원 소개

← ↻ 🔒 https://meet.google.com/bmu-hrgb-enh

m minju Kim님이 발표 중입니다.

23/01/14

Yamuzin Notion

🔍 검색

🕒 업데이트

⚙️ 설정과 멤버

> 🇰🇷 삼성카드 데이터 분석 &...

> ⭐ Tensorflow Study

> 🌟 CUI AI D팀

> 🏆 Future Finance AI Chall...

개인 페이지

> 📅 2023

> 🌐 2022 CUI AI

> 📅 2학기 스터디

> 📅 At Ellice

> 📅 At EY

> 🌟 Yamuzin_oksusu

> 🌟 민주의 복잡한 세상살이

> 📅 데이터 분석 스터디

📁 팀스페이스 생성

🔔 알림

📁 가져오기

+ 새 페이지

< > + 🌐 2022 CUI AI / ... / 📁 Meeting Log / 4 23/01/14

4

23/01/14

- Electra
 - 여러 개를 마스킹 → 하나씩 추론해서 list형식으로 나옴
- Konlpy → 형태소 분석
 - 📄 조사_패딩.ipynb 746.9KB
- T5
 - 🔗 kolang-t5-base
 - 토큰화를 따로 해서 사용할 수 있으면 되지 않을까

meet.google.com에서 내 화면을 공유하는 중입니다. 공유 중지 숨기기

공유

연결하기

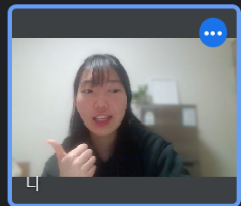
이 Notion 페이지에 액세스하는 모든 사용자에게 실시간으로 업데이트되는 미리보기가 표시됩니다. 자세히 알아보기

연결

2:58 오후 | bmu-hrgb-enh

🔊 📹 🗨️ 🖱️ 😊 📶 📶 📶 📶 📶

이혜연 (응용통계 21) 김민주 (경영 18)



김민기 (경영 17) 김지민 (응용통계 20)

주제 확정

브로카 실어증을 가진 환자의 언어 치료를 위한 양방향 대화형 모델 구축

문장을 생략하고 문법에 맞지 않는 문장으로 의사소통을 하며, 간단하게 표현하는 증상

데이터셋 : AI Hub 한국어 데이터셋 활용

 AI Hub

AI 데이터찾기 AI 개발지원 참여하기 정보공유 고객센터 AI 허브소개 로그인 회원가입 

데이터 분야  | AI 데이터찾기 > 데이터 분야

데이터셋 검색 

객체별 검색 데이터 유형별 검색

분야 선택

데이터유형 선택

한국어

영상이미지

헬스케어

재난안전환경

농축수산

교통물류

 이미지

 비디오

 텍스트

 오디오

 3D

 센서

초기화 

MLM 모델 탐색

환자의 발화

나 밥 먹다



MLM 모델 input

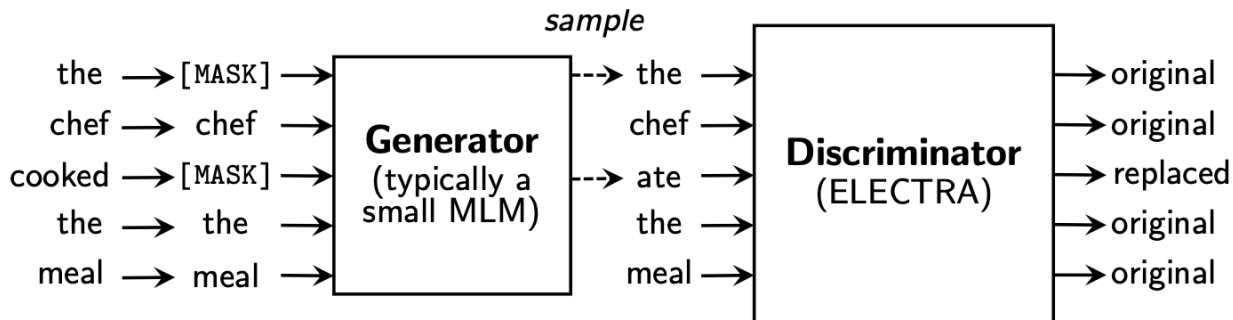
나<Masked> 밥<Masked> 먹<Masked>



MLM 모델 output

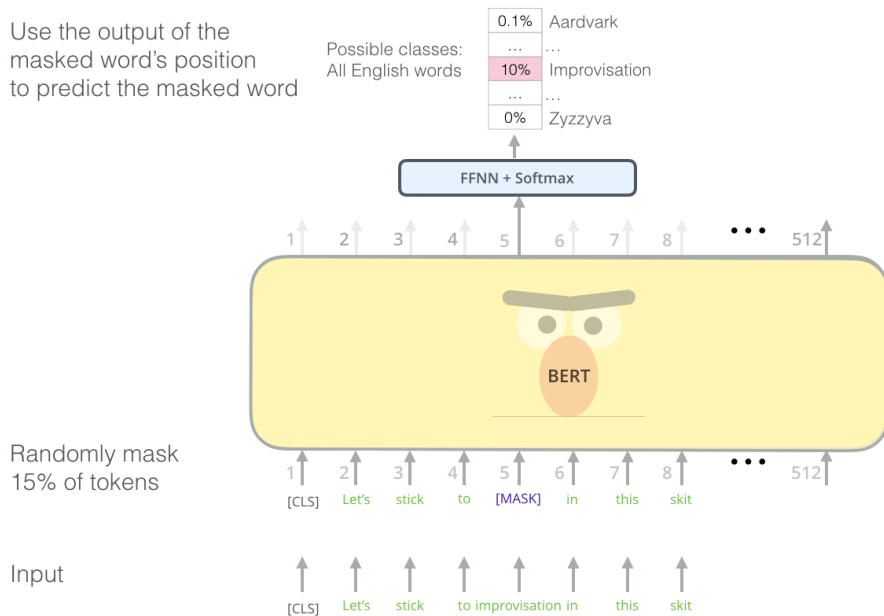
나는 밥을 먹었어요.

1) ELECTRA 모델 -> 한국어 TEXT로 학습된 koELECTRA 모델



2) Konlpy 형태소 분석 -> BERT 모델

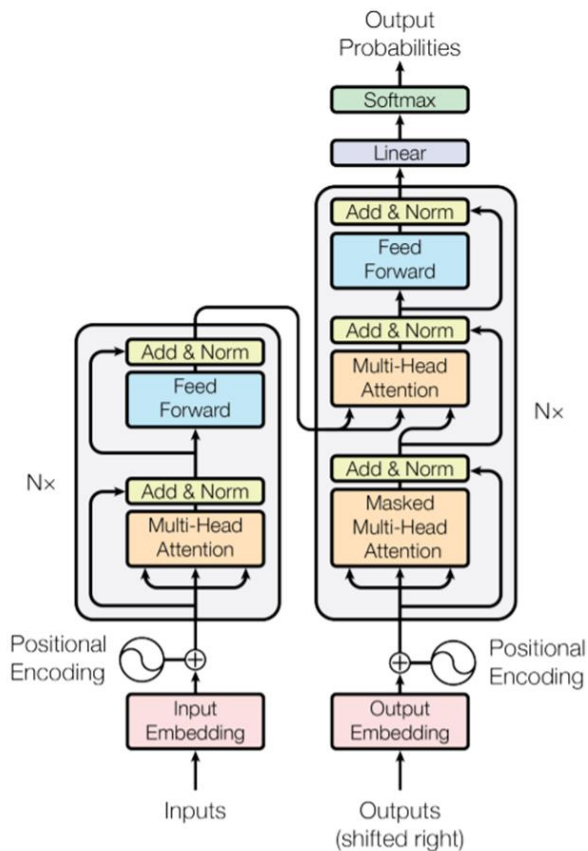
Use the output of the masked word's position to predict the masked word



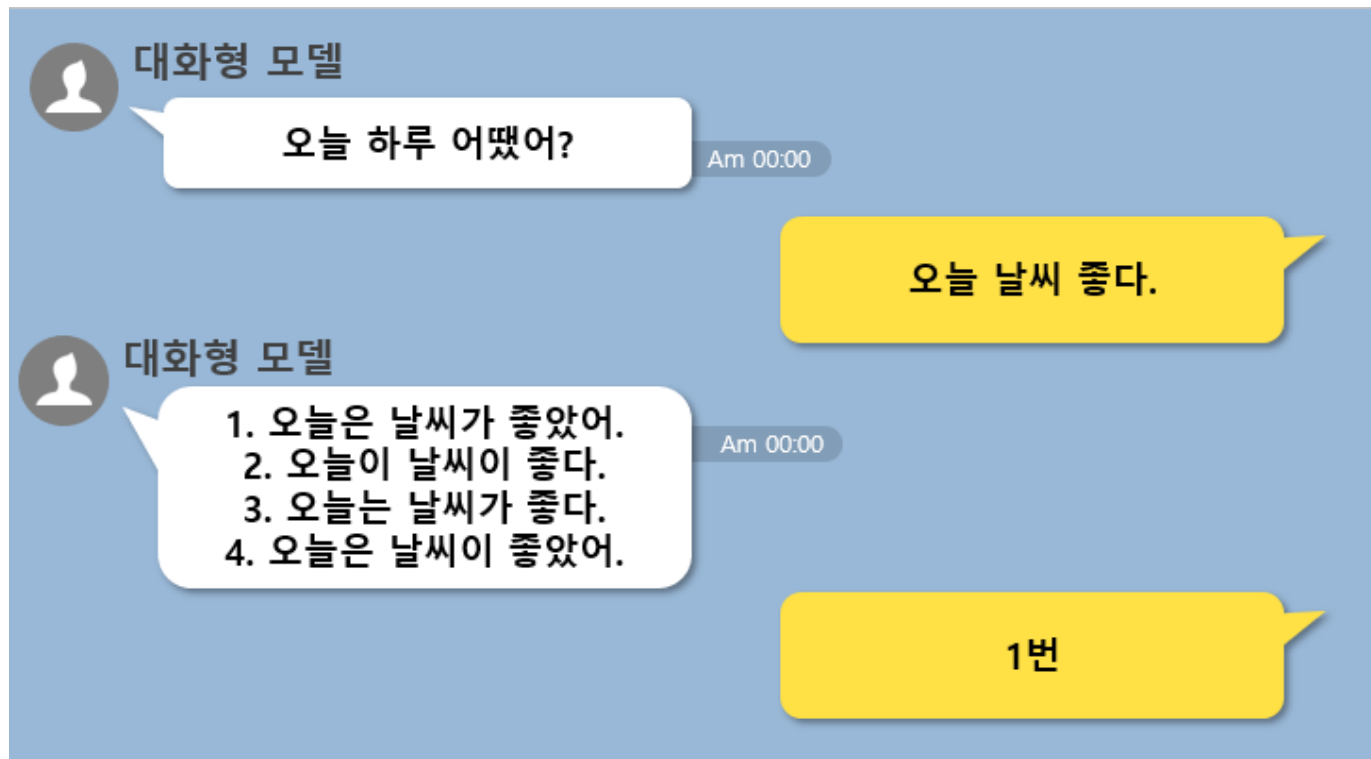
**BERT의 encoder, GPT의 decoder layer를
합친 T5 아키텍처 생성 ->**

3) T5 모델

(Text-to-Text Transfer Transformer)



양방향 대화형 치료 모델



이후 방향

1) 모델 프로토타입 조사 및 학습 (T5 모델, 조사 패딩 부분)

-> 발화에 맞게 올바른 문장 생성

2) 챗봇 또는 다른 방식을 조사 및 학습

-> 환자와 양방향으로 소통할 수 있게 구현

=> 양방향 소통이 가능한 대화형 치료 모델 구축

감사합니다!

CUAI 동계 컨퍼런스 NLP팀