**Weekly report**

날짜 : 2022-07-28

연구원 : 최윤석

* 수행결과
* 트라콤에서 준 데이터를 학습할 수 있게 전처리 작업
  + SSD에서 학습할 수 있게 코드를 만들어 각각 라벨에서 200개의 이미지와 라벨 추출
* BERT의 임베딩 방식과 BERT에서 어텐션 메커니즘을 어떻게 사용하는지 학습
  + 포지셔널 임베딩, MLM, NSP와 BERT의 인코더 구조와 멀티헤드 셀프어텐션 학습
* 윤종찬 연구원께서 논문 이해를 위해 추천해주신 RELBERT 관련 학습
  + 셀프 어텐션을 활용해 레코드를 BERT를 활용해 자연어 문장처럼 임베딩
* 관계형 테이블 임베딩에서 BERT-base가 아닌 large를 사용하는 방안
  + 윤종찬 연구원의 의견으로는 412호 서버 하드웨어가 좋지 않아 large 모델이 안될 것 같음
* 관계형 테이블 임베딩에서 숫자형 데이터를 추가하는 방향
  + 실수형 데이터를 임베딩할 수 있는가?
    - 0.12548 같이 특정한 데이터를 전부 임베딩할 수 없음 -> 실수 데이터를 정규화하는 방향
* 수행계획
* 추가적인 전처리 작업 이후 트라콤 데이터를 활용해 SSD 모듈 생성 및 추론
* BERT에서 실수형 데이터를 정규화 / 이산화해서 임베딩하는 방법
* 기타사항

▶ 트라콤 미팅 8/8일 오후 1시 30분