**Weekly report**

날짜 : **2023-06-08**

연구원 : **최윤석**

* **이전 수행계획**
  + **관계형 테이블 임베딩 :**
    - Named Entity Recognition할 문장을 만들기 위해 Column name + is + Column value 구조에서 추가적으로 덧붙여 문장 내 단어의 수를 늘릴 방법 고민 (메타데이터 활용 등..)
    - 수업시간에 발표한 범주형 – 범주형 컬럼간의 조인 아이디어를 구현한 코드 작성
* **수행결과**
* **관계형 테이블 임베딩** : (임베딩 / 융합 가능성 파트 분류해서 작성).
  + 수업시간에 발표한 범주형 – 범주형 컬럼간의 조인 아이디어를 구현한 코드 작성
    - Column name embedding -> pre-trained FastText 모델을 사용
    - Named Entity Recognition -> BERT를 이용하여 Kaggle에 있는 영어 NER Data를 학습 데이터로 하여 모델 학습
    - Containment -> 직접 함수를 작성
    - 텍스트, 폰트, 화이트, 스크린샷이(가) 표시된 사진

      자동 생성된 설명35개의 테이블 데이터들을 입력으로 모델을 실행한 결과
    - 조인 가능한 테이블쌍과 어느 컬럼을 기준으로 조인이 가능한지를 출력해줌
* **수행계획**
  + **관계형 테이블 임베딩 :**
    - Named Entity Recognition할 문장을 만들기 위해 Column name + is + Column value 구조에서 추가적으로 덧붙여 문장 내 단어의 수를 늘릴 방법 고민 (메타데이터 활용 등..)
    - 성능평가 기준에 맞게 데이터셋에 라벨링을 하여 성능평가 진행, 이후에는 기존의 다른 모델들과 성능 비교
* **기타사항**
  + **리포트 작성 시, 논문의 내용정리 같이 남이 한 부분에 대한 설명은 자세하게 하지 말고 그 중에서 내가 얻은 아이디어나 내 연구에 관한 부분을 자세하게 설명할 것**
  + **테이블 간 조인 / 유니온 하기위한 컬럼을 찾는 문제에 관해 생각 (Elmo / Bert / Word2Vec) 등도 생각 / 내가 손으로 하는 것을 자동화하는 방법으로 생각**
  + [**https://koreaai.org/**](https://koreaai.org/) **<- 한국인공지능학술대회**
  + [**http://bic-conference.org/2023/**](http://bic-conference.org/2023/) **<- BIC학술대회 (7/1까지 제출)**