**Weekly report**

날짜 : **2023-06-16**

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..)
    - 내가 만든 모델의 성능평가 기준을 세우고, 성능평가 기준에 맞게 데이터셋에 라벨링을 하여 성능평가 진행, 이후에는 기존의 다른 모델들과 성능 비교
    - 데이터셋 개수 늘려서 모델 실험 진행
* **수행결과**
* **관계형 테이블 임베딩** : (임베딩 / 융합 가능성 파트 분류해서 작성).
  + 데이터셋의 수를 35개에서 50개로 늘려서 모델 실행, csv 파일을 읽는 코드에서 발생하는 오류 해결
    - 50개의 테이블 데이터들을 입력으로 모델을 실행한 결과

텍스트, 폰트, 흑백, 스크린샷이(가) 표시된 사진

자동 생성된 설명

* + - 조인 가능한 테이블쌍과 어느 컬럼을 기준으로 조인이 가능한지를 출력해줌
    - 지금까지 읽은 조인가능한 테이블들을 찾는 논문들의 성능평가 방법 중 가장 좋은 아이디어를 나의 성능평가에 사용
      * Precision 계산 : 내 모델이 찾은 조인 가능한 컬럼들 중 실제로 조인이 가능한 컬럼들 / 내 모델이 찾은 조인 가능한 컬럼들
      * Recall 계산 : 내 모델이 찾은 조인 가능한 컬럼들 중 실제로 조인이 가능한 컬럼들 / 실제로 조인이 가능한 컬럼들
        + 실제로 조인이 가능한 컬럼들을 찾는 방법

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

* + - * + 비교할 다른 Methods들을 모델로 만들고 각 모델에 실험에 사용할 테이블 데이터셋을 넣어줌
        + 각 모델들이 테이블 데이터셋에서 조인이 가능한 컬럼들을 출력하면 그 결과들을 전부 UNION 연산을 진행함.
        + UNION 연산을 진행한 조인 가능한 컬럼들을 실제로 조인이 가능한 컬럼이라고 생각함
        + 이 방법의 장점 : 테이블 데이터를 하나하나 살펴보면서 어느 컬럼들이 조인 가능한지 직접 찾아내는 수고를 덜 수 있어, 성능평가시 사용할 테이블 수를 늘리는 데 부담이 줄어듦.
* **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
    - [**http://bic-conference.org/2023/**](http://bic-conference.org/2023/) **<- BIC학술대회 (7/1까지 제출)**
* **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**