**Weekly report**

날짜 : **2023-11-22**

연구원 : **최윤석**

* **이전 수행계획**
  + **관계형 테이블 임베딩 :**
    - CNE-join의 성능을 개선할 방법 고민
      * 조인하는 테이블의 개수를 3개 이상으로 확장해보기, 최적의 테이블 개수는 무엇인가? 이를 확인할 척도에 대해서도 고민
        + CNE-join의 결과로 나온 2개의 조인 가능한 테이블쌍 중에서 A-B, B-C 같이 테이블이 겹치는 쌍을 활용해 n-join-pair를 생성하는 알고리즘 및 이에 대한 융합도 산정 방안 고민 및 소스코드 작성(이전 CNE-join과의 일관성이 중요함)
      * 시간 단축 문제는 가장 나중으로 진행
    - Measuring and Predicting the Quality of a Join for Data Discovery 논문이 참고한 문헌을 읽고 n-join-pair 알고리즘 아이디어 얻기
    - **CNE-join 결과로 나온 조인 가능한 테이블쌍 정보를 활용해 최대 4개의 테이블을 조인한 결과를 벡터로 임베딩하여 2차원 공간에 시각화하기**
    - 박찬훈 연구원과 DataMap 개발(Type A, B)
    - **고급정보검색론 수업시간에 발표한 논문 보충해서 추가로 작성하기**
      * **임베딩과 텍스트 분류 사이의 연결고리**
      * **논문에서 중요한 부분 중 빠트린 부분**
* **수행결과**
  + **관계형 테이블 임베딩 :**
  + 조인 테이블 개수를 4개 이하로 테이블을 임베딩한 벡터들을 2차원 공간에 시각화하기
  + 윤종찬 연구원의 제안모델인 TabBERT를 일부분 수정하고 사용해, 하나의 테이블을 하나의 벡터로 임베딩함

텍스트, 포스트잇 노트, 폰트, 스크린샷이(가) 표시된 사진

자동 생성된 설명

Figure . 윤종찬 연구원의 TabBERT 제안모델

* + 윤종찬 연구원의 코드를 수정하고, 이를 이용해 CNE-join에 사용한 50여개의 테이블들을 각각 하나의 벡터로 임베딩함.
  + CNE-join의 결과로 나온 조인 가능한 테이블쌍을 활용해 두 테이블을 조인하여 만든 조인 테이블들도 각각 하나의 벡터로 임베딩함.

텍스트, 폰트, 스크린샷, 문서이(가) 표시된 사진

자동 생성된 설명

그림 . CNE-join에 사용한 50여개의 테이블들과 조인한 테이블들을 2차원에 시각화한 그림

* CNE-join 결과로 나온 조인 가능한 테이블쌍 정보를 활용해 3개, 4개의 테이블도 조인하여 2차원 공간에 시각화할 예정
* 고급정보검색론 수업시간에 발표할 논문 보충해서 추가로 PPT 작성
* **수행계획**
  + **관계형 테이블 임베딩 :**
    - CNE-join의 성능을 개선할 방법 고민
      * 조인하는 테이블의 개수를 3개 이상으로 확장해보기, 최적의 테이블 개수는 무엇인가? 이를 확인할 척도에 대해서도 고민
        + CNE-join의 결과로 나온 2개의 조인 가능한 테이블쌍 중에서 A-B, B-C 같이 테이블이 겹치는 쌍을 활용해 n-join-pair를 생성하는 알고리즘 및 이에 대한 융합도 산정 방안 고민 및 소스코드 작성(이전 CNE-join과의 일관성이 중요함)
      * 시간 단축 문제는 가장 나중으로 진행
    - Measuring and Predicting the Quality of a Join for Data Discovery 논문이 참고한 문헌을 읽고 n-join-pair 알고리즘 아이디어 얻기
    - CNE-join 결과로 나온 조인 가능한 테이블쌍 정보를 활용해 최대 4개의 테이블을 조인한 결과를 벡터로 임베딩하여 2차원 공간에 시각화하기
      * 2개 테이블 조인한 것, 3개 테이블 조인한 것, 4개 테이블 조인한 것을 각각 다른 그림에 넣어서 구별이 쉽게 시각화하기(테이블이 너무 많아서 시각화 했을 때 보기 힘듦
      * 시각화한 결과에서 얻은 insight를 뒷받침할 근거나 멘트 생각하기
    - 박찬훈 연구원과 DataMap 개발(Type A, B)
* **기타사항**