**Weekly report**

날짜 : **2023-10-11**

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
        + 이를 위해 2개의 sequence들이 있을 때 이를 n개의 sequence로 확장시키는 알고리즘에 대해 공부
      * 시간 단축 문제는 가장 나중으로 진행
    - 박찬훈 연구원과 DataMap 개발(Type A, B)

* **수행결과**
  + **관계형 테이블 임베딩 :**
* CNE-join 결과로 나온 테이블쌍 중에서 겹치게 나오는 하나의 테이블을 중심 테이블로 하여 최대 n개까지의 테이블을 조인함.

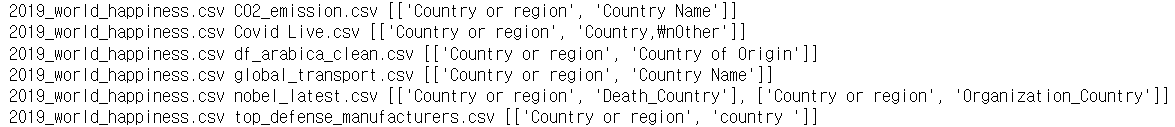


Figure 1. CNE-join 실행 결과 중 2019\_world\_happiness 테이블과 조인이 되는 테이블

텍스트, 번호, 스크린샷, 평행이(가) 표시된 사진

자동 생성된 설명

Figure 2. 2019\_world\_happiness 테이블을 중심으로 조인

* 조인 방식은 일단 INNER JOIN을 사용할 것이고, 다른 조인도 사용해서 가장 좋은 조인 방식을 사용할 예정임.
* 아래의 식을 만족하면 조인이 잘 되었다고 판단할 것임
  + > Threshold

텍스트, 폰트, 스크린샷, 화이트이(가) 표시된 사진

자동 생성된 설명

* + 추가로 조인이 잘 되었다고 판단하는 기준으로 n-join-pair를 만들 때, 중심 테이블이 아닌 주변 테이블들 테이블 임베딩하여 하나의 벡터로 만들고, 테이블 벡터간 코사인 유사도 계산을 통해 값이 높으면 조인이 잘 되었다고 판단할 예정임  
    테이블 **A**, B, C, D
  + A-B, A-C, A-D
  + B,C,D -> 임베딩하여 Vector B,C,D
* **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
    - CNE-join 특허출원 관련 최항구 변리사님과 미팅(10/17 화요일 오후 2시 화상미팅)
    - 박찬훈 연구원과 DataMap 개발(Type A, B)
* **기타사항**