**Weekly report**

날짜 : **2024-05-17**

연구원 : **최윤석**

* **이전 수행계획**
  + **관계형 테이블 임베딩 :**
    - DataMap 개발(Type A, B)
    - 자동 융합 시나리오에 대한 명세를 작성해보기(그림으로) – 논문 서론에 사용할 용도로 간략화 추상화된 그림을 작성
      * 쿼리 바이 테이블도 설명은 하되, 나는 어디까지 구현을 했는지를 표현
      * 서비스 사용자 입장에서 어떻게 융합된 테이블을 얻게 되는지
      * 서비스 개발자 입장에서 무엇을 하는지
      * PPT로 작성 Type A B 구분하지 말고
      * 융합을 할 때 Join을 할 지 Union을 할 지 확인하는 방법 – 나는 Join 파트
    - 테이블 임베딩을 위한 Feature selection 방법 탐색
      * 테이블에서 임베딩 해야 하는 중요 컬럼을 자동 / 휴리스틱을 이용해 알아내는 방법 탐색
      * 나만의 Feature selection 아이디어가 있으면 연구의 Sub결과로 활용 가능하니 아이디어의 변형이 있으면 고민해보기 - CNE-Join에서 Feature selection으로 시간 줄이기 가능 / N-Join에서 Feature selection은 성능 향상으로 가능
    - TF-IDF 방식의 평가 척도 개선
      * 예시를 통해 논리를 보강할 필요가 있음, 예시를 우선적으로 제시하고 TF-IDF의 일반적인 방식을 소개하는 것으로
      * 아이디어가 너무 심플해 보임. 살짝은 더 로직을 추가
        + 재료 테이블 TF-IDF 1위부터 10위 중 1위에 가까운 단어들이 조인 테이블 TF-IDF 30위 안에 들어 있을 시 더 높은 가중치를 두는 방식 -> 해당 방식으로 실험하기
      * 튜닝 아이디어 : 재료 테이블과 융합 테이블에 가중치를 부여하는 방법?
        + 재료 테이블의 전체 레코드 수 와 융합 테이블의 전체 레코드 수가 조인 품질 척도의 인자로 들어가는 것도 생각해보기
    - 본 심사 준비
      * 학위논문 수정
        + 피드백 내용 반영

데이터셋에 설명에 대한 보완 / 기존의 기법이 쓴 데이터셋을 활용

테이블이 많아졌을 때 사용이 힘듦. complexity study를 통해 이론적으로 가능한 실험인지 확인

비교 기법이 기존 기법 중 어떤 것 인지 달기

데이터베이스 쪽 용어를 사용하기보단 다른 용어를 사용하는 방안

라벨링 과정중에 애매한 것들을 직접 설명하면서 설득력 얻기

CNE-Join 실험결과가 압도적으로 좋은 이유에 대한 설명 필요

* **수행결과**
  + **관계형 테이블 임베딩 :**
    - 본 심사 준비
      * 피드백 받은 내용 및 각각의 피드백에 대한 대응방안을 PPT로 정리
    - 빅데이터딥러닝 수업시간에 발표할 논문 읽고 PPT 작성
    - 객프 중간고사 채점
* **수행계획**
  + **관계형 테이블 임베딩 :**
    - DataMap 개발(Type A, B)
    - 자동 융합 시나리오에 대한 명세를 작성해보기(그림으로) – 논문 서론에 사용할 용도로 간략화 추상화된 그림을 작성
      * 쿼리 바이 테이블도 설명은 하되, 나는 어디까지 구현을 했는지를 표현
      * 서비스 사용자 입장에서 어떻게 융합된 테이블을 얻게 되는지
      * 서비스 개발자 입장에서 무엇을 하는지
      * PPT로 작성 Type A B 구분하지 말고
      * 융합을 할 때 Join을 할 지 Union을 할 지 확인하는 방법 – 나는 Join 파트
    - 테이블 임베딩을 위한 Feature selection 방법 탐색
      * 테이블에서 임베딩 해야 하는 중요 컬럼을 자동 / 휴리스틱을 이용해 알아내는 방법 탐색
      * 나만의 Feature selection 아이디어가 있으면 연구의 Sub결과로 활용 가능하니 아이디어의 변형이 있으면 고민해보기 - CNE-Join에서 Feature selection으로 시간 줄이기 가능 / N-Join에서 Feature selection은 성능 향상으로 가능
    - TF-IDF 방식의 평가 척도 개선
      * 예시를 통해 논리를 보강할 필요가 있음, 예시를 우선적으로 제시하고 TF-IDF의 일반적인 방식을 소개하는 것으로
      * 아이디어가 너무 심플해 보임. 살짝은 더 로직을 추가
        + 재료 테이블 TF-IDF 1위부터 10위 중 1위에 가까운 단어들이 조인 테이블 TF-IDF 30위 안에 들어 있을 시 더 높은 가중치를 두는 방식 -> 해당 방식으로 실험하기
      * 튜닝 아이디어 : 재료 테이블과 융합 테이블에 가중치를 부여하는 방법?
        + 재료 테이블의 전체 레코드 수 와 융합 테이블의 전체 레코드 수가 조인 품질 척도의 인자로 들어가는 것도 생각해보기
    - 본 심사 준비
      * CNE-Join 데이터 개수 1000개정도로 늘려서 실험
      * 학위논문 수정
        + 피드백 내용 반영
        + 실험 파트에 데이터맵을 case study로 내용 추가
* **기타사항**
  + **6월 저널 논문 투고**