**Weekly report**

날짜 : **2022-09-16**

연구원 : **최윤석**

* **수행결과**
* **차량 이미지 분류 :** 모델 학습시 발생하는 오류의 해결중
* **관계형 테이블 임베딩** :
  + 캐글에서 텍스트 데이터와 수치형 데이터가 혼합된 테이블 데이터셋(혼합형 데이터셋)을 수집 (현재 총 32개 데이터)

텍스트이(가) 표시된 사진

자동 생성된 설명

Figure 1. 혼합형 데이터셋 모음



Figure 2. 혼합형 데이터셋 예시

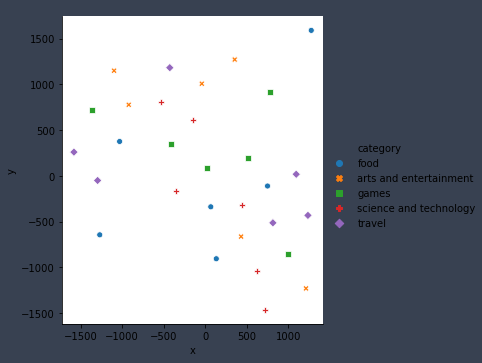
* + 단순 계급 값 이산화 (1-5사이 값으로 이산화), 이도훈 연구원의 방식으로 이산화 (int(), 단순 계급 값 이산화한 것을 범주형 데이터로 (very high(5) / high(4) / medium(3) / low(2) / very low(1)) 변경하는 3가지 방식의 코드를 만들고 비교실험   
    

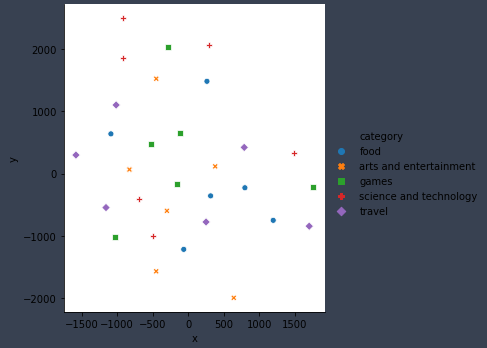
Figure 3. 단순 계급 값 이산화  


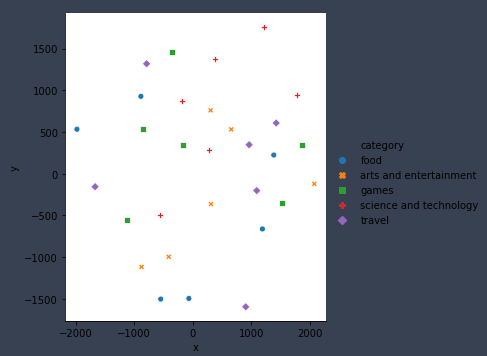
Figure 4. 이도훈 연구원의 방식으로 이산화  


Figure 5. 이산화한 것을 범주형 데이터로 변경

* + 혼합형 테이블 데이터를 이산화하고, 윤종찬 연구원의 기존 코드를 일부 수정하여 테이블을 임베딩
* **수행계획**
* **차량 이미지 분류** : 학습 도중 발생하는 오류 해결 및 모델 학습
* **관계형 테이블 임베딩** :
  + 지속적으로 카테고리에 부합하는 데이터셋을 추가 (목표 전체 데이터셋 150개 이상)
  + 추가한 데이터셋을 이용해 비교실험을 다시 진행
  + NDCG 계산에 필요한 Dataset\_rel.csv 파일 생성 및 NDCG로 임베딩 모델의 정량적 평가 진행
* **기타사항**
  + **데이터 융합 프로세스**

