서론 설문조사 항목 작성 (채현) -> 수요일 설문조사지 fix

본론 내용 작성, gbm 구현 (미섭)

분류 데이터 EDA 및 시각화 (정아)

<보고서 목차>

1. 서론

주제 선정 이유 (서비스 필요성) - 현황

기능 소개

[본론]

1. 데이터(수집 & EDA)
   1. 동아리 탐색 키워드 데이터
   2. 동아리 선호 평점 데이터
   3. 동아리 특징 설문 데이터
   4. 동아리 소개 책자 데이터
2. 모델구현 (원리->모델링->결과(분석))
   1. 키워드를 이용한 동아리 탐색(SQL)
   2. 동아리 유사도 기반 동아리 추천(contents-based)
   3. 동아리 선호 평점 기반 동아리 추천(item-based)
   4. 분과별 2021 동아리 트렌드 (text mining)
   5. 동아리 선호 평점 기반 동아리 추천(collaborative filtering)

SVD, SVD++, NMF

앙상블 방법(수치평균, 회귀, gbm)

* 1. 개인 특성 & 동아리 정보 기반 동아리 추천

1. 결론
   1. 향후 과제
      1. 애기능 동아리, 단과대 동아리, 연합 동아리
      2. 웹페이지에서 동아리 특징 설문 데이터 수집

<보고서 평가>

아이디어 : 창의성, 적합성, 적용성, 활용성, 현실성

데이터 : 구현성, 활용성, 시의성, 적합성

{한계}

많은 사용자들로부터 얻은 기호 정보를 통해 추천 시스템을 가동하는 ‘협업 필터링’은 만족스러운 결과를 도출할 수도 있지만 단점도 존재합니다. 먼저 ‘콜드 스타트(cold start)’라고 일컬어지는 문제입니다. 협업 필터링 알고리즘을 사용하기 위해서는 기존 데이터가 반드시 필요하기에 아무런 행동이 기록되지 않은 신규 사용자에게는 어떠한 아이템도 추천할 수 없게 됩니다. 콜드 스타트란 이러한 상황을 일컫는 말로 ‘새로 시작할 때의 곤란함’을 의미합니다.

두 번째는 계산 효율 저하입니다. 협업 필터링은 계산량이 비교적 많은 알고리즘이기 때문에 사용자 수가 많은 경우 계산이 몇 시간에서 길게는 며칠까지 소요되기도 합니다.

마지막으로 롱테일(long tail) 문제입니다. 사용자들은 소수의 인기 있는 콘텐츠에만 관심을 보이기 마련이고, 자연스럽게 소수의 인기 콘텐츠가 전체 추천 콘텐츠의 많은 비율을 차지하게 됩니다. 인기가 저조한 콘텐츠는 정보 부족으로 추천되지 못하는 셈이죠.