3주차 회의록

- 협업필터링 모델 추가 구현 (Item-based)

- 분류모델 추가 구현 (Bagging)

- 협업필터링 추천시스템 평가 지표에 대한 코드 구현

- 전체 사용자에 대한 추천 결과 데이터 구축

- 평가 지표 산출(Personalization, Intra Similarity)

- 설문조사 대상자에 따른 설문조사 분류와 세분화

- 동아리 정보 수집을 위한 동아리연합회 정기회의 참여

- 동아리연합회로부터 동아리 박람회 때 사용한 각 동아리 공식 홍보자료 워드 파일 수령

- 동아리 현 활동 기수로부터 동아리 설문조사 시행

- 웹 UI 설계

[이번 주에 한 일] ~목요일까지

* *설문조사 대상자에 따른 설문조사 분류, 세분화*
* *동아리 키워드 정리*
* *동아리연합회 정기회의 발표자료 준비(제공하는 서비스, 기대효과 정리, 설문지캡쳐)- 정아채현 [데이터 얻기 마지막!]*
* *인프런 강의 수강,, 서버 더 공부하기 - 이정아씨….*
* 설문지 직접 만들기<순서도 정렬하기> -미썹
* *UI 설계 - 나머지도!*
* *디지털 정보처 협력 메일 보내기 -채*
* 모델 조금 더 손봐서 완성시키기 ! -미썹
* Collaborative Filtering 추천시스템 평가 지표에 대한 코드 구현

전체 사용자에 대한 추천 데이터 구축,

평가 지표 산출 (Personalization, Intra Similarity)

<모델>

추천 알고리즘 Item-based 기반 협업 필터링 추가

분류 알고리즘 앙상블 모델 추가

Collaborative Filtering 추천시스템 평가 지표에 대한 코드 구현

<데이터>

설문조사 대상자에 따른 설문조사 분류와 세분화

동아리 키워드 재정리

동아리 정보 수집을 위한 동아리연합회 정기회의 참여 및 발표 준비

<웹>

UI 설계 및 백엔드 구현 과정 탐색

[해야할 일]

* surprise library의 SVD 함수 가상환경에서 import 가능한지
* contents based code 업데이트 (정아) O

<필요성>

* 기존의 동아리 정보는 동아리연합회 책자, 에브리타임&고파스의 홍보게시판에서 열람 가능
* 에브리타임&고파스의 경우 동아리의 정보를 알리기보단 홍보 목적(for 지원자 모집)이며 중복 게시물이 많아 정보 열람 어려움 -> 결과적으로 새로운 동아리를 알기보단 기존에 알고 있는 동아리를 검색하는 정도
* 동아리 책자를 온라인으로 구현해 검색 기능 구현 (검색과 추천이 자유롭다)

<서비스>

* 사용자 평점에 따른 동아리 추천
* 사용자 특성에 따른 동아리 추천
* 키워드별 동아리 검색
* 동아리 검색
* 동아리 정보를 포함한 특징 제공
* 가입 동시에 동아리 평점 매기기

<동연에게 제공>

* 동아리별 특징 제공 (설문조사 결과 시각화)
* 동아리 키워드 정리한 파일

<기대효과> (서비스 상용화)

* 동아리 홍보 효과, 활성화
* 정확한 동아리 특징 파악 가능 (설문 데이터 쌓임)
* 동아리 홍보 자료 첨부, 정보 업데이트 가능
* 동아리 연합회 홈페이지의 미흡함을 보완할 수 있는 사이트로서의 가치 어필

[총 이번주에 한거]

* 데이터 수집 최종, 모델 구현 최종, 서버 시작 , ui 디자인 시작

[앞으로 할거 다음주,,, 4주에,,, ]

* 보고서 부분 : 데이터 수집, 모델구현(설문조사 분류모델 제외) 부분 작성
* 개발 집중 : DB(키워드, 검색창, 평점 기반, 설문조사 학습데이터, 사용자용 응답데이터, 회원정보데이터), 응답 서버 고르기(백엔드 마무리)
* UI 최종 디자인 완료
* (미섭) NLP - Contents-based & CF(Word2Vec 등) + 주석

[아마도 5주 쯤,,,]

- 데이터 확보 완료! 분류모델 마무리

- 보고서 부분 : 데이터 수집, 모델 구현 최종 작성

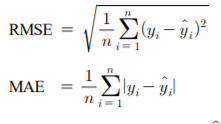
- 개발 : 백엔드 마무리하고 프론트 공부해서 키워드랑 검색창, 설문조사 학습데이터 정도 구현

[아마도 6주쯤..]

* 보고서 최종 작성
* 개발 프론트 마무리 + 모든 기능 구현 + 배포까지 해보기!

---------------------------------------------------------------------------------------------------------------------------

(mae 조사 결과)



어느 지표가 더 좋은지에 대한 객관적인 사실은 없다. MAE의 경우 비교 수준의 차이가 있을 때 사용하기 적절하지 않다. (키의 경우 160~180, 몸무게의 경우 50~80일 때 후자의 경우가 MAE가 더 작게 나올 가능성이 크다.) 하지만 영화 평점의 경우 모두 같은 5 scale이기 때문에 MAE의 활용이 가능하다는 점에서 사용해도 좋을듯 !?

다른 논문에서 MAE, RMSE 두개를 주로 써서 정확도 지표를 보더라. 하나의 평가 기준만 사용하지 않는듯해

(만약 hyperparameter를 구하려면 하나를 지정해야 하는데 에러에 대한 가중치를 어떻게 잡고 싶느냐에 따라 달라질듯?)

→ 난 MAE가 평균적으로 몇점의 평균 오차가 있는지 볼 수 있어서 넣는것도 좋을거같엉

MAE

- 실제 값과 예측 값의 차이(Error)를 절대값으로 변환해 평균화

- MAE는 에러에 절대값을 취하기 때문에 에러의 크기 그대로 반영된다. 그러므로 예측 결과물의 에러가 10이 나온 것이 5로 나온 것보다 2배가 나쁜 도메인에서 쓰기 적합한 계산식이다.

- 에러에 따른 손실이 선형적으로 올라갈 때 적합하다.

- 이상치가 많을 때

RMSE

- MSE 값은 오류의 제곱을 구하므로 실제 오류 평균보다 더 커지는 특성이 있어 MSE에 루트를 씌운 RMSE 값을 쓴다.

- 에러에 제곱을 하기 때문에 에러가 크면 클수록 그에 따른 가중치가 높이 반영된다. 그러므로 예측 결과물의 에러가 10이 나온 것이 5로 나온 것보다, 정확히 2^2(4)배가 나쁜 도메인에서 쓰기 적합한 계산식이다.

- 에러에 따른 손실이 기하 급수적으로 올라가는 상황에서 쓰기 적합하다.