



Wrap-up Report

RecSys_3조 (ㄱ해쥬)

강성택, 김다빈, 김윤경, 김희수, 노근서, 박영균

1. 프로젝트 개요

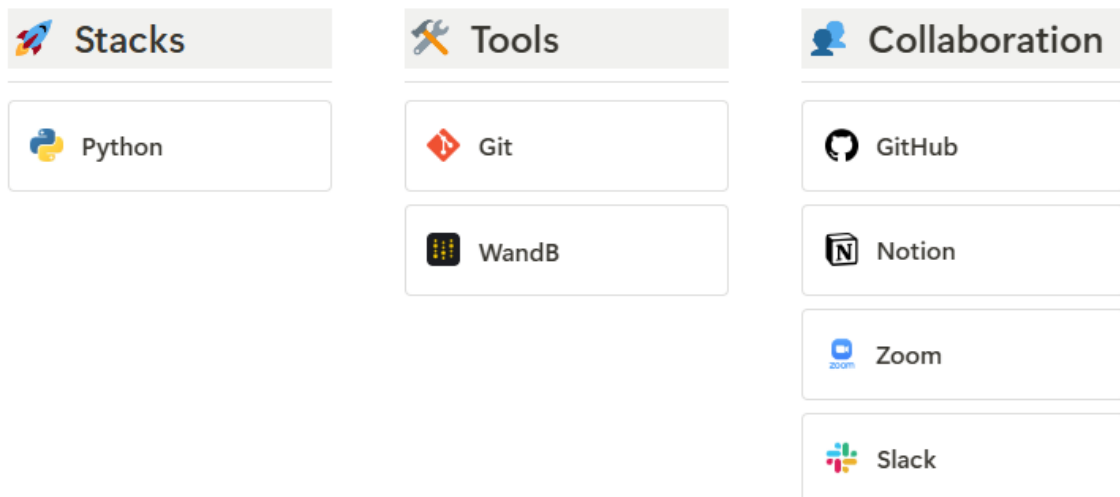
1.1 개요

사용자의 영화 시청 이력 데이터 셋(사용자의 Implicit 데이터, 영화의 meta 데이터)을 활용하여 각 사용자가 다음에 시청할 영화 및 좋아할 영화에 대해 예측한다.

이는 곧 소비자들의 영화 시청 결정을 주기 위한 개인화된 상품 추천과 같다.

1.2 환경

- (팀 구성) 6인 1팀
- (컴퓨팅 환경) V100 서버를 SSH로 연결하여 사용
- (협업 환경) Notion, Github, WandB
- (의사 소통) Zoom, Slack



2. 프로젝트 팀 구성 및 역할

데이터팀 / 모델팀으로 3:3 나눈 후, 각 팀마다 한 명씩 짝지어서 **페어 프로그래밍** 방식으로 진행

공통

서버 구축, EDA

Data 팀

- 공통 - 전처리 및 피처 엔지니어링 모듈화
- 노근서 - replace_duplication, merge_dataset, id2idx, df2mat 구현, MultiVAE
- 김희수 - multi_hot_encoding, 파생변수 생성 함수들 구현
- 박영균 - filter_top_k_by_count, fill_na, preprocess_title 구현, BERT4Rec 구현

Model 팀

- 공통 - 모듈화, 코드 리팩토링
- 김윤경 - Baseline 작성, RecBole 구현
- 김다빈 - DeepFM, Baseline 작성, WandB, 모델 파이프라인 구축
- 강성택 - EASE, EASER, MultiVAE, Recbole, CatBoost 구현 및 실험

3. 프로젝트 수행 절차 및 방법

3.1 팀 목표 설정

진행 기간 : 11/13 (수) 10:00 ~ 11/28 (목) 19:00

- (1주차) 강의 전부 듣기, 데이터셋 이해, 각자 EDA 완성
- (2주차) 데이터 전처리, 모델 모듈화 및 실험
- (3주차) 모델 실험 및 앙상블

3.2 프로젝트 계획

1. Project Rule, Github/Coding Convention(PEP8 기반) 설정
2. 프로젝트 개발 환경 구축 - 서버 세팅, Github Template 적용
3. 데이터셋 사전 공부
4. EDA 진행 및 인사이트 도출
5. 데이터팀/모델팀 코드 개발 & 모듈화
6. EDA별 실험 및 성능 비교(WandB 사용) & 모델 튜닝 : 교차 검증, 하이퍼파라미터 튜닝
7. 앙상블
8. 코드 리팩토링 & 프로젝트 코드 최종 모듈화

3.3 협업 방식

- 이슈 관리 - Github의 Issues + Notion 데이터베이스 보드로 진행
 - Github의 Issues : Issue Template(개요, TODO 작성)
 - 총괄 보드 : 전체 진행 상황 관리(일정, 제출 관리 등)
 - Data 보드 : 데이터 전처리, 피처 엔지니어링 함수 및 모듈화 진행상황 관리
 - Model 보드 : 모델 설계, 모듈화, 앙상블 개발 등 진행상황 관리
- Github Convention에 따라 Github 관리
- Coding Convention에 따라 개발

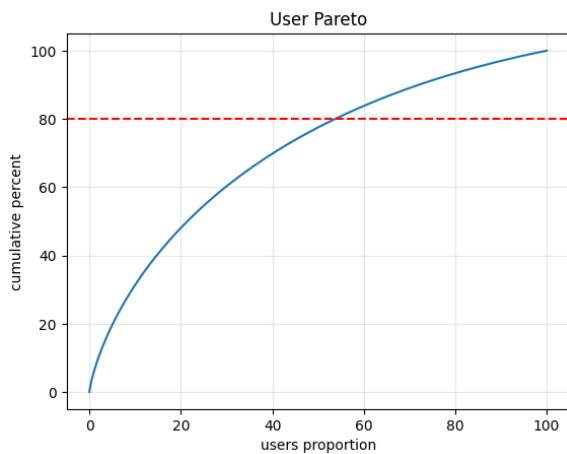
- WandB를 활용한 실험 관리

4. 프로젝트 수행 결과

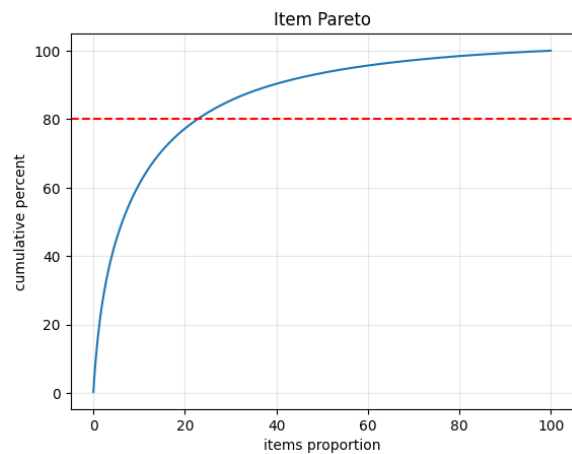
4.1 탐색적 데이터 분석(EDA)

누락된 연도 (title 마지막에서 추출)	item	제목	원제	감독
1902	32898	달세계 여행	A Trip to the Moon	조르주 멜리에스
1915	7065	국가의 탄생	Birth of a Nation	D. W. 그리피스
1916	7243	인톨러런스	Intolerance: Love's Struggle Throughout the Ages	D. W. 그리피스
1919	6988	흩어진 꽃잎	Broken Blossoms or The Yellow Man and the Girl	D. W. 그리피스
1917	8511	이민자	The Immigrant	찰리 채플린
1921	3310	키드	The Kid	찰리 채플린
1920	6987	칼리가리 박사의 밀실	The Cabinet of Dr. Caligari	로베르트 비네
2015	119145	킹스맨	Kingsman: The Secret Service	매튜 본

year 변수에 누락된 아이템이 8개 존재함.

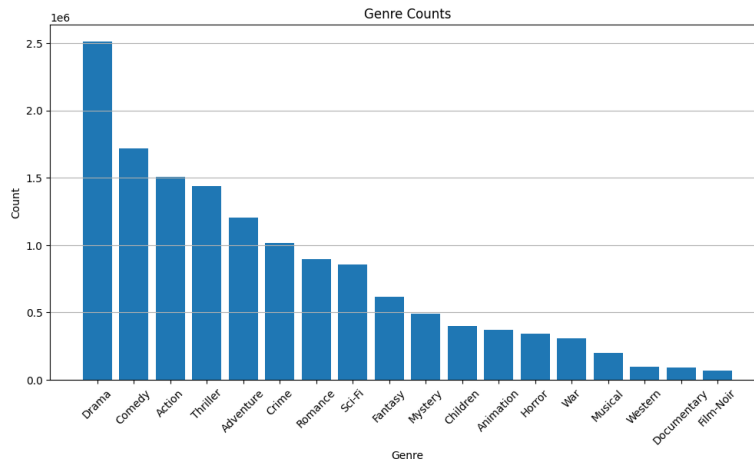


유저 롱테일 특성 확인



아이템 롱테일 특성 확인

유저에 대해서는 파레토 법칙이 성립하지 않지만, 아이템에 대해서는 파레토 법칙이 성립함



장르별 리뷰 수 분포

장르별로 리뷰의 수가 크게 차이남. 따라서 아래의 가설을 세움

user별로 선호하는 genre가 있고 해당 genre의 영화를 많이 본다

더 나아가, 선호하는 genre와 연관된 genre의 영화도 많이 본다

4.2 데이터 전처리

- `year`: 결측치를 title에 있는 연도 정보로 채우기
- `title`의 영문제목에서 특수문자를 제거한 뒤, 소문자로 변환하여 사용
- `director`: 하나의 영화에 감독이 여러 명인 경우, 데이터 전체 영화에 대한 감독의 빈도가 높은 순으로 2명까지 사용, 또한 item에 대해 director가 결측치가 있으면 -1로 대체
- `genre`: 하나의 영화에 장르가 여러 개인 경우, 데이터 전체 영화에 대한 장르의 빈도가 높은 순으로 4개까지 사용
- `item`: item 34048과 64997은 title, year, writer, genre는 전부 같지만, 64997의 경우는 director만 결측치이다. 따라서 item 64997을 34048로 대체

4.3 피쳐 엔지니어링

[파생변수 생성]

- 영화 리뷰수(`num_reviews_item`): 영화의 리뷰 횟수로 영화의 인기 정도를 나타내는 피쳐를 추가
- 리뷰 유무(`review`): negative sampling을 위한 implicit feedback 피쳐를 추가
- 장르 취향(`genre_taste`): 유저별 장르의 취향을 파악하기 위한 피쳐를 추가
- 감독 취향(`director_taste`): 유저별 감독의 취향을 파악하기 위한 피쳐를 추가

[변수 선택]

- 범주형 변수: `user`, `item`, `director`, `genre`, `book_title`, `review`, `genre_taste`, `director_taste`
- 수치형 변수: `num_reviews_item`, `year`, `time`

4.4 모델 선정 및 평가

[모델 선정 및 이유]

• General

EASE

- **데이터 특성에 적합:** 데이터셋의 복잡도가 낮고, 사용자-아이템 간 상호작용이 명확하여 간단한 행렬 연산 기반 접근이 효과적일 것으로 예상
- **효율성:** Sparse한 데이터에서 매우 빠르고 효율적으로 작동하며, 학습 및 추론 속도가 빠름
- **모델 구조:** Shallow Autoencoder(AE) 기반으로, 저차원(low-order) 패턴을 학습하여 데이터의 주요 특징을 포착
- **Sparse 데이터에 최적화:** 데이터 희소성을 고려해 Gram 행렬 기반 접근으로 아이템 간의 관계를 효과적으로 학습

EASER

- **EASE의 확장 모델:** EASE 구조에 추가적으로 Smoothing 기법을 적용해 데이터 불균형 문제를 완화하여 EASE보다 높은 성능을 보일 것이라 예상
- **안정성:** 특정 아이템이나 사용자의 높은 가중치에 민감한 EASE의 한계를 개선해 일반화 성능을 높임
- **희소 데이터 개선 효과:** Smoothing을 통해 가중치가 지나치게 집중되지 않도록 균형을 맞추어 안정적인 성능 제공

MultiVAE

- **Sparse 데이터 학습:** 희소 데이터에서도 Variational Autoencoder 구조를 통해 잠재 공간(latent space)을 학습하며 패턴을 효과적으로 포착
- **복잡한 패턴 탐지:** 단순한 저차원 패턴(EASE/EASER) 외에 고차원(high-order) 패턴과 비선형 관계를 학습하여 높은 성능을 보일 것이라 예상
- **개인화 추천 강화:** 사용자와 아이템 간의 잠재적 관계를 학습하여 개인화 추천에 강점이 있음

• Sequential

SASRec

- **시간적 연속성 모델링:** 시청 이력의 시간적 순서를 반영해 다음에 시청할 아이템을 예측하는 Transformer 기반 Sequential 모델
- **Self-Attention 활용:** 사용자 행동에서 중요한 정보에 가중치를 부여해 연속적인 상호작용 패턴을 학습
- **시계열 데이터 적합:** 시간적 패턴을 효과적으로 학습하며, Sequential 추천에서 높은 성능을 보일 것이라 예상

• Context

DeepFM

- **Feature Interaction 학습:** 사용자와 아이템의 상호작용을 학습하며, 선형 관계(FM)와 비선형 관계(Deep Neural Network)를 모두 학습

- **메타데이터 활용에 강점:** Sparse Feature와 Dense Feature를 모두 처리할 수 있어 사용자/아이템의 다양한 메타데이터를 활용 가능
- **복잡한 관계 학습:** 비선형적이고 복잡한 데이터 패턴도 포착하여 높은 추천 성능을 보일 것이라 예상

• Tree

CatBoost


- **빠른 학습 및 예측:** GPU 학습을 지원해 대규모 데이터에서도 빠르게 학습 가능
- **모델 해석 가능성:** Feature Importance를 제공해 모델의 결과를 해석할 수 있다는 점에서 도입

[모델 평가]

다양한 모델(EASE, EASER, MultiVAE, SASRec, DeepFM, CatBoost)을 활용해 **Recall@10**, **NDCG@10** 성능을 비교

4.6 모델 성능 및 결과

리더 보드[중간 순위]

8	RecSys_03조		0.1603	33	2d
---	------------	--	--------	----	----

리더 보드[최종 순위]

8	RecSys_03조		0.1602	33	3d
---	------------	---	--------	----	----

제출 목록

<input checked="" type="checkbox"/>	easer_0.0001		0.1603 0.1602	2024.11.26 23:51	완료	
제출1. EASER						
<input checked="" type="checkbox"/>	EASE_test		0.1600 0.1602	2024.11.25 14:47	완료	
제출2. EASE						

4.7 시행 착오

[RecBole]

다양한 모델 실험을 진행하기 위해 RecBole을 구현하면서 다음과 같은 과정을 거쳤다

`create_dataset` 하는 과정에서 RecBole에서 요구하는 `.inter` 파일이 자동으로 생성되지 않아 데이터를 RecBole 형식에 맞게 전처리해야 했다. 이에 따라 `user` 와 `item` ID를 고유 숫자로 매핑하고, `user_id:token` 과 같은 형식으로 변환한 후 탭(`\t`) 구분자로 저장해 `.inter` 파일을 생성했다.

이후 모델 학습을 바탕으로 topk를 추출하는 과정에서 다양한 시행착오를 겪었다.

1. `inference.py` 에서 저장된 모델을 불러오기 위해 직접 저장된 파일의 이름을 적어야 했다. 하지만 저장된 모델 파일이 많아질수록 불편함을 느꼈고, 저장된 모델 파일 중 가장 최근 파일을 찾는 `find_recent_model` 함수를 생성하여 이 과정을 자동화하였다.
2. RecBole에서 제공하는 함수인 `full_sort_topk` 함수에서 `history_item` 필드가 자동으로 생성되지 않는 문제가 발생하였다. 이는 모델 간 차이로 인해 `history_item` 이 항상 포함되지 않아 생긴 문제로, 제공된 함수를 사용하지 않고 사용자와 이미 상호작용한 아이템을 직접 계산한 후, 이 아이템들을 추천 목록에서 제외하는 방식으로 수정했다.

하지만 사용자와 아이템 간 점수를 직접 계산하는 과정에서 매우 많은 시간이 소요되었다. 이는 모든 사용자의 상호작용 기록을 순회하며 추천 목록에서 필터링하는 방식으로 구현했기 때문에 발생한 성능 병목이었다.

이로 인해 대규모 데이터셋에서의 성능 병목을 방지하기 위한 방법에 대해 고려할 필요가 있었다는 점을 깨달았고, 효율적인 데이터 파이프라인 구축의 필요성을 몸소 느끼게 되었다.

[다중 범주 데이터 처리]

장르와 감독에 대한 전처리를 진행하던 중, 여러 범주를 가지는 데이터에 대해 어떻게 나타내는 것이 적절한지 고민이 되었다.

1. 여러 범주를 각각의 행으로 분리해 하나의 사용자-아이템 조합에 대해 여러 행을 만드는 방식
2. 여러 범주를 배열 형태로 묶어 그대로 변수로 할당하는 방식

1번 방식으로 데이터를 생성하면 하나의 사용자-아이템 조합에 대해 여러 개의 예측 결과가 주어지는 문제가 발생하고, 2번 방식을 이용할 경우 모델에서 이를 적절히 처리할 지에 대한 근본적인 의문이 들었다.

이에 대해 서치해본 결과는 다음과 같다:

여러 범주를 행 방향으로 복사하는 방식

- 각 범주의 특징을 독립적으로 고려할 수 있으며, 주로 전통적인 머신러닝 기법을 활용할 때 유용하다.
- 최종 예측 점수를 산출할 때에는 예측 점수의 평균값, 최빈값, 최댓값 등의 후처리를 이용하면 문제가 없다.

배열 형태로 직접 변수에 할당하는 방식

- 다양한 범주 정보를 한꺼번에 학습하므로 복잡한 모델(딥러닝 기반 모델 등)을 이용할 때에 유용하다.
- 그러나 모델이 기본적으로 배열 형태의 입력을 처리할 수 있는 구조인지 파악해야 한다.

위의 과정을 거치며 데이터 구조가 모델 학습에 미치는 영향에 대해 깊이 고민해볼 수 있었다.

5. 자체 평가 의견

잘한 점

- **다양한 모델 구현** : 대회 개요와 EDA를 바탕으로 각자가 생각하는 적절한 모델과 그 이유에 대해 상의하며 구현할 모델들을 선정하였다. 이에 따라 모델의 구조적 특성과 데이터의 관계를 깊이 이해하고, 데이터 특성에 적합한 모델 전략을 설계할 수 있었다.

- **끝까지 최선을 다하는 태도:** 마지막까지 모두 열심히 몰입하는 모습이 인상적이었다. 비록 계획이 틀어지고 후반부에 목표를 수정하는 등 우여곡절이 많았지만, 끝까지 프로젝트에 최선을 다하는 팀원들의 태도가 돋보였다.
- **적극적인 피드백 반영:** 지난 프로젝트의 피드백을 바탕으로 협업, 기록, 코드 구현 등 다양한 방식에서 발전했다. 특히 WandB를 활용해 실험 관리와 시각화를 체계적으로 수행했고, 각자의 역할을 세분화해 소통하여 효율적으로 진행할 수 있었다. 결과적으로 협업 방식이 더 원활해졌다.

아쉬운 점

- **구현능력 & 모델지식 부족:** 프로젝트 진행 중 일부 구현에서 예상보다 많은 시간이 소요되었고, 특정 모델의 원리나 특성에 대한 이해가 부족해 어려움을 겪었다. 특히, 새로운 기법을 적용하거나 모델 성능을 개선하는 과정에서 세부적인 조정에 시간이 많이 걸렸는데, 이는 관련 모델 지식이 더 보완되어야 할 필요성을 느꼈다.
- **데드라인 관리:** 프로젝트가 시작되고 본격적으로 진행하기 전에 팀원들끼리 일정을 미리 작성하였음에도 불구하고 일정이 제대로 지켜지지 않아 아쉬움이 남는다.
- **컨디션 관리:** 이번 프로젝트는 저번 프로젝트와 달리 시간이 넉넉하게 주어졌음에도 불구하고 팀원들이 만족하는 프로젝트가 이루어지지 않았다. 그 이유는 프로젝트가 길었던 만큼 다들 밤늦게까지 일을 하였기 때문에 컨디션을 관리하기 힘들었기 때문이다. 이리하여 컨디션 관리를 조절하지 못해 번아웃이 와버려 프로젝트 마지막 즈음에 다들 프로젝트에 대한 집중이 흐트러진 부분에 대해 아쉬움이 남는다.

이번 프로젝트를 통해 배운 점

→ 여전히 3 GUNA!

1. 모델 구현을 위해 모델 공부는 필수구나!
2. 프로젝트 과정 속에서 몸 관리 및 멘탈 관리는 필수구나!
3. 그리고 언제나 도메인 지식은 알면 알수록 중요하구나!

다음 프로젝트에 적용해볼 점

→ 3 HAGA!

1. 이번 프로젝트 구조를 참고하여 다음 프로젝트 진행하자!
2. 가장 성능이 뛰어난 모델을 기준으로만 실험을 진행하는 것이 아닌 다양한 모델과 함께 실험을 진행하면서 기록하자!
3. 모델을 단순히 점수 기준으로만 사용하지 말고, 모델 구조에 대한 이해를 한 후, 모델을 선정하자!

6. 개인 회고

강성택_T7501

1. 나의 학습목표

다양한 추천 알고리즘(EASE, EASER, MultiVAE, SASRec 등)을 학습하고, 실제 데이터를 통해 구현 및 성능 평가를 경험하는 것이 목표였다. 특히 Sparse 데이터와 Sequential 데이터에서 효과적인 추천 모델을 설계하고, 모델 간의 성능 차이를 이해하는 데 중점을 두었다.

2. 전과 비교해서, 내가 새롭게 시도한 변화는 무엇이고, 어떤 효과가 있었는가?

딥러닝 기반 추천 모델(MultiVAE, SASRec)을 기존 행렬 연산 기반 모델(EASE, EASER)과 비교하며 학습을

하였다. 이를 통해 모델별 장단점을 명확히 이해할 수 있었고, 데이터 특성과 요구 사항에 따라 적합한 모델을 선택하는 능력을 기를 수 있었다.

2. 개인 학습 측면

프로젝트를 통해 Sparse 데이터 처리, Sequential 모델링, 딥러닝 기반 추천 시스템에 대한 실질적인 구현능력을 기를 수 있었다. 특히, 하이퍼파라미터 튜닝 및 모델별 성능 비교를 통해 각 모델의 내부 동작 원리를 깊이 이해할 수 있었다.

3. 공동 학습 측면

팀원들과의 협업을 통해 각 모델의 구현과 평가를 분담하며 효율적으로 작업을 진행했었다. 모델 성능을 비교하고, 개선 아이디어를 도출하는 과정에서 추천 시스템 설계에 대한 다양한 관점을 배울 수 있었다. 다만 똑같은 오류가 발생하였을 때, 해결방법이 공유되지 않아 각자 개별적으로 해결하였었는데 이를 통해 소통의 중요성을 깨달았다.

4. 기억에 남는 시도

먼저 대회에 데이터와 개요를 보았을 때, 접근방법에 대해 깊게 고찰하는 시간을 가졌었다. 도메인 지식을 활용하여 영화를 선택할 때 주로 장르, 감독 등 주요 요소를 고려한다는 점에서 복잡한 관계보다는 단순한 상호작용 패턴이 더 중요한 영향을 미칠 것이라 예상했었고, 주어진 데이터를 분석하였을 때 Sparse한 데이터 특성을 가졌었다. 이를 종합하여 Sparse한 데이터 특성과 low-order 패턴을 학습하는 데 뛰어난 EASE를 도입하고자 하였고, 모든 모델 중 가장 뛰어난 성능을 보였다.

5. 한계와 아쉬움

EASE를 기반으로 추천을 생성한 뒤, high-order 패턴을 학습하는 모델과 앙상블하면 더 나은 성능을 기대할 수 있을 것이라는 생각을 하였었다. 하지만, RecBole을 시도하는 과정에서 오류가 많이 발생하였고, 시간이 지연되어 결국 시도해보지 못한 점이 아쉬웠다.

6. 앞으로의 도전

이번 프로젝트의 목표는 다양한 모델을 다뤄보는 것이었다. 그동안 부족했던 모델에 대한 지식과 구현 능력을 기를 수 있었고, 이 지식과 경험을 잃지 않고 다음 프로젝트에서 적극 활용하고 싶다.

김다빈_T7558

1. 나의 학습목표

이번 프로젝트에서 모델팀으로서 최소한 모델 하나는 구조를 파악하고 직접 구현하는 것, 그리고 모듈화를 통해 코드의 재사용성과 유지보수성을 높이는 것이 목표였다.

2. 전과 비교해서, 내가 새롭게 시도한 변화는 무엇이고, 어떤 효과가 있었는가?

이전 프로젝트들에서는 다소 수동적인 태도로 주어진 역할에만 충실한 경향이 있었다. 하지만 이번에는 주어진 작업 외에도 팀원들과 적극적으로 소통하며 아이디어를 제안하고, 프로젝트의 전반적인 방향성에 관여하려고 노력했다. 맡은 일에서만 책임감을 가지는 게 아닌, 프로젝트 전체에 대한 책임감을 가지는 계기가 되었다.

2. 개인 학습 측면

모델 구현에 필요한 자료를 찾을 때, 다양한 학습 자료와 문서를 참고했지만, 처음에는 어떤 자료를 선택해야 할지에 대한 판단이 어려웠다. 여러 자료를 참고하면서, 실제로 필요한 정보와 예시 코드를 찾는 법을 배웠다. 앞으로는 참고한 자료를 따로 정리해 팀원들과도 공유하고 이후에 효율적으로 활용하고 싶다.

3. 공동 학습 측면

모델팀끼리 초기 구조를 만들 때 오류를 겪었을 때 각자만의 방식으로 시도를 해보고 해결책을 공유했다. 이런 과정을 통해 처음에는 막막했던 부분들이 명확해졌고, 놓치는 부분 없이 꼼꼼히 점검할 수 있었다.

처음에는 효율적이고 유지보수 가능한 코드 구조를 목표로 했지만, 쉽게 이해하고 사용할 수 있는 구조로 구현하지 못했다. 앞으로는 명확하고 간단한 모듈화가 이루어지도록 구조를 개선하고, 팀 내에서 코드 흐름을 쉽게 이해할 수 있도록 문서화하는 작업이 필요하다는 점을 배웠다.

4. 기억에 남는 시도

DeepFM 모델은 입력 데이터의 각 샘플에 대해 단일 예측값을 출력한다. 이런 특성을 가진 모델에 입력한 시퀀스에 대해 전체 아이템 집합에 대한 예측값을 출력하도록 되어있는 Sequential 모델 베이스라인 코드를 참고해서 적용해 어려움을 겪었다. 이 과정에서 모델의 다양한 특성을 알았고, 각 특성에 맞춘 구현이 필요하다는 것을 배웠다.

5. 한계와 아쉬움

모듈화를 통해 구조화된 코드를 작성하고 이에 맞추어 원하는 모델을 구현하고자 했다. 이를 위해 코드의 각 단계를 명확히 분리하고 재사용 가능한 구성 요소로 만들기 위해 노력했다. 그러나 구조화 과정 자체에서 예상보다 많은 시간과 노력이 소요되면서, 정작 구현하고자 했던 모델은 끝내 시도조차 하지 못했다는 점이 가장 아쉽다. 또한, 구조화 과정에 너무 몰입한 나머지 핵심 목표인 모델 구현 및 실험에 대한 집중도가 떨어졌다는 점도 아쉽다.

6. 앞으로의 도전

다음 프로젝트는 마지막 프로젝트이며 동시에 데이터 전처리 및 모델 구현에서 더 나아가 배포까지 다룰 수 있는 중요한 기회이다. 때문에 일련의 과정에서 담당하고자 하는 부분에 대해 미리 공부해서 활용하고 싶다.

김윤경_T7511

1. 나의 학습목표는 무엇이었나?

이번 프로젝트에서는 지금까지 배운 추천 모델을 더 깊이 있게 공부하고 실제 데이터에 적용해보는 것이 목표였다. 특히 논문을 읽고 해당 모델을 구현까지 해보면서 앞으로의 모델 공부에서 더 중점적으로 다뤄야 할 부분을 스스로 체감해보는 데에 중점을 두었다.

2. 전과 비교해서, 내가 새롭게 시도한 변화는 무엇이고, 어떤 효과가 있었는가?

전에 모델 팀을 담당했을 땐 feature engineering 한 데이터를 실험해보기 위한 틀을 마련한다는 느낌이 강했다면 이번에는 모델 공부를 바탕으로 직접 적용해보는 시간이었다. 물론 성공하진 못했지만 덕분에 모델 공부에 좀더 신경써야 한다는 점을 더욱 확실히 느꼈다.

3. 개인 학습 측면

오류를 마주했을 때 문제 해결 접근 방식의 중요성을 배웠다. 이전에는 오류가 발생하면 해당 코드에만 집중하며 해결하려고 했지만, 이번 경험을 통해 코드의 전체적인 흐름을 살펴보는 것이 더 효과적일 수 있다는 점을 깨달았다. 또한 문제 해결이 너무 오래 끌게 되면 전체적인 프로젝트 진행에 영향을 미칠 수 있다는 점에서 데드라인을 설정하고 적절히 시간 관리를 해야 한다는 교훈도 얻었다.

4. 공동 학습 측면

지난 프로젝트부터 이어온 협업 방식이 효율적으로 개선되어 체계적인 틀로 자리잡은 것 같다. 그러나 팀원 모두 프로젝트 진행 중 컨디션 관리에 어려움을 겪으면서 피로가 누적되었고, 이로 인해 예상보다 효율이 떨어진 순간들도 있었다. 앞으로는 협업의 효율성을 유지하기 위해 체계적인 방식뿐만 아니라 팀원들의 컨디션 관리에도 더 신경 써야 한다는 점을 배웠다.

5. 한계와 아쉬움

RecBole 라이브러리를 활용하면서 자동으로 생성되어야 할 몇몇 파일들이 예상대로 생성되지 않는 상황을 마주해 난항을 겪었다. 이 과정에서 다양한 방법으로 문제를 해결하려고 노력했지만, 완전히 해결하지 못한 부분이 남아 있어 아쉬움이 있다. 앞으로는 좀더 다양한 방식으로 해결하려 고민해보고 개선해 나가야겠다고 느꼈다.

6. 앞으로의 도전

앞으로는 팀원들과 함께 효율성과 성과를 모두 잡는 프로젝트를 만들어 가고 싶다. 체력과 집중력이 프로젝트 전체의 효율성과 결과에 큰 영향을 미친다는 점과 내가 생각한 것보다 모델 공부는 더 깊이 있게 해야한다는 점을 이번 경험을 통해 깊이 깨달았다. 따라서 프로젝트의 전체적인 계획을 세워 컨디션 관리와 모델 공부, 프로젝트 진행까지 모든 면에서 한 단계 성장하고자 한다.

김희수_T7513

1. 나는 내 학습목표를 달성하기 위해 무엇을 어떻게 했는가?

저의 학습목표는 추천시스템 모델에 대하여 공부하는 것이었다. 이를 위해 책이나 논문, GPT 등을 사용하여 공부하였다.

2. 전과 비교해서, 내가 새롭게 시도한 변화는 무엇이고, 어떤 효과가 있었는가?

전과 비교해 새롭게 시도한 것은 모델에 대해 이해하려는 것이었다. 이해하기 위해 여러 가지 참고사항들을 보았지만 너무 겉핥기 식으로 이해하여 프로젝트에 있어 큰 영향을 발휘하지 못했다.

3. 마주한 한계는 무엇이며, 아쉬웠던 점은 무엇인가?

제 주관적으로 생각하였을 때 이번 프로젝트에 있어 핵심은 모델이라고 생각한다. 하지만 저의 빈약한 모델 지식으로 인하여 이번 프로젝트에 있어 팀원들에게 도울 수 없을 뿐만 아니라 팀원들의 발목을 잡았던 것 같다. 그나마 팀원들에게 도움이 되기 위해 영화감독에 대한 결측치를 채우려 하였지만 이마저도 실행하지 않아 아쉬움이 남는다. 그리고 컨디션관리를 실패하여 번아웃이 와버려 프로젝트에 성실하지 않은 태도를 보인 것이 아쉽다.

4. 한계/교훈을 바탕으로 다음 프로젝트에서 시도해보고 싶은 점은 무엇인가?

다음 프로젝트는 이제 이 팀원들과 하는 마지막 프로젝트이다. 이때까지 받은 피드백, 제가 생각한 문제점, 예로 GitHub, 모델 지식, 코딩 구현, 데드라인 약속, 컨디션 관리 등등을 다 지키면서 라스트 댄스를 춰보고 싶다.

노근서_T7514

1. 나의 학습 목표는 무엇이었나? 달성하기 위해 무엇을 어떻게 했는가?

본격적인 프로젝트를 시작하기 전에 어떤 모델을 도입하면 프로젝트에 더 효과적일지 미리 고민하고 이를 구현까지 옮길 수 있도록 실천해보는 것이 목표였고, 목표를 실천하기 위해서 논문을 포함한 모델 공부가 필수적이라고 생각했다. (결과적으로 뜻대로 되진 않았지만..)

2. 전과 비교해서, 내가 새롭게 시도한 변화는 무엇이고, 어떤 효과가 있었는가?

데이터팀에서 모델팀에서 필요로 하는 작업을 서포팅하는 것을 넘어 적용시키고 싶은 모델까지 직접 실험해보는 과정을 겪었다. 한 프로젝트에서 전반적인 파이프라인을 직접 체험해보면서 각 팀에서의 작업 과정에 어떤 고민들이 있는지 몸소 느낄 수 있던 경험을 할 수 있어서 좋았다.

3. 마주한 한계는 무엇이며 아쉬웠던 점은 무엇인가?

미리 계획을 잡아두고 프로젝트를 진행했음에도 불구하고 중간 과정에 진척이 생기니 전체 프로젝트 스케줄에 영향을 미친다는 걸 크게 느낄 수 있었다. 어쩌면 이전 프로젝트들에선 이 정도로 작업 과정에서 꼬이는 일이 없다보니 겪은 성장통이라고 받아들이니 좋은 경험이라 생각한다.

4. 한계/교훈을 바탕으로 다음 프로젝트에서 시도해보고 싶은 점은 무엇인가?

스케줄 관리를 지금과는 다른 방법으로 도입해보는 것도 좋을 듯 하다. 기존까지는 줌으로 회의한 내용을 노션에 TODO 관리와 함께 적는 방법을 활용했다. 어느 정도 안정권에 들어섰다는 생각에 비슷한 방식의 작업 과정이 반복되었으니, 마침 멘토링 시간에 추천도 받았고 여러 팀들도 사용한다는 Jira 또는 Linear와 같은 협업 툴을 새롭게 사용해보면서 새로운 시도를 경험하는 것도 좋다고 생각한다.

5. 내가 해본 시도 중 어떤 실패를 경험했는가? 실패 과정에서 어떤 교훈을 얻었는가?

데이터 구조가 계층 형태일 때와 배열 형태일 때 단순히 데이터 구조만으로도 모델이 학습할 때 변수 사이의 관계를 다르게 인식하는 지 궁금해 실험을 해보려고 시도했다. 두 가지 구조에 모두 적용되는 함수를 구현하지 못해 결국 따로 적용하는 방법을 고려할 수 밖에 없었고 전처리, 피쳐 엔지니어링 과정도 2배로 걸렸다. 모델팀 작업도 쉽지 않아 결국 하고 싶었던 실험은 못한 채 모델팀 작업을 도우러 갔다. 너무 모든 가능한 상황에 맞는 코드로 일반화 시키려고 했던 점이 오랜 작업 시간으로 나타났다고 생각해 어떻게 하면 효율적인 프로젝트 시간 관리를 할 수 있을지 다시 한 번 깨달을 수 있는 시간이었다.

박영균_T7520

1. 나의 학습목표는 무엇이었나?

모델의 전체적인 구조를 이해하는 것과, 이를 넘어 코드 상으로 구현할 수 있는 것을 목표로 하였다. 하지만 너무 완벽하게 하려는 생각이 앞서 어느 것도 제대로 하지는 못한 듯하다.

2. 전과 비교해서, 내가 새롭게 시도한 변화는 무엇이고, 어떤 효과가 있었는가?

이전 세 개의 프로젝트는 모델 파트를 도맡아 했지만, 이번에는 데이터 전처리 및 피쳐 엔지니어링 역할을 시도했다. 예상한 대로 모델 파트에 흥미가 있다는 사실을 다시 알게 된 것도 있지만, 그 외에 나의 강점과 흥미에 대해 확실히 알게 되었다는 점에서 의의가 있었던 것 같다.

3. 개인적으로 어떤 것을 배웠는가?

프로젝트 내용 외적인 것이지만, 계획대로 되지 않는 일도 충분히 많이 벌어지며, 그 다음 어떻게 하는지가 더 중요하다는 것을 알았다. 지난 프로젝트와 달리 제대로 되는 것들이 없어 매우 힘들었는데, 멘토님과의 상담에서 위의 내용을 배웠다. 내가 할 수 있는 선에서 최선을 다하면 그걸로 된 것이라는 생각을 항상 가지면서 임해야겠다.

4. 협업 측면에서 얻은 것이 있는가?

협업에서는 크게 달라진 것이 없어서 그런지 잘 모르겠지만, 다른 팀원이 말하기를 프로젝트 기간에 너무 눈치를 보느라 제때 쉬러 가지 못하는 것 같다는 말을 듣고 느낀 것이 있었다. 프로젝트를 수행하는데 있어 컨디션 관리가 실력만큼 중요하고, 그게 가능하도록 팀원 서로간의 노력이 뒷받침되어야 할 것 같다.

5. 프로젝트 중 기억에 남는 것이 있는가?

이번 프로젝트의 PM역할을 맡았는데, 크게 한 일은 없지만 왠지 나쁘지 않게 역할을 해냈다는 생각이 들었다. 그리고 나서 알게 되었는데, 장이라는 직책에 너무 의미부여해서 부담을 갖지는 않아도 된다는 점이였다.(그렇다고 너무 소홀해지면 안될 것 같기도 하다..)

6. 프로젝트에서 아쉬웠던 점은?

아무래도 건강관리가 시급하다. 운동도 하지 않고 달려오다보니 체력이 부족해서 집중력이 흐려지는게 보일 정도였다. 지속가능한 학습을 위해 쉬어가는 겸 틈틈이 운동을 해야겠다.

7. 다음 프로젝트에서 시도해볼 것은?

항상 미뤄왔던 논문 보기를 이번 기회에 습관화해야겠다. 정독이 아니더라도 핵심 내용을 파악하고 이를 나의 모델 구현에 적용할 수 있는 능력을 기르고 싶다.