

부스트캠프 AI Tech 7기 기업 해커톤 (TVING)

Recsys-7조(X10)

곽정무, 박준하, 박태지, 배현우, 신경호, 이효준

1. 프로젝트 개요

1-1. 프로젝트 개요

해당 프로젝트는 부스트캠프 AI Tech 7기 기업 해커톤 중 TVING 주제 2에 해당하는 주제인 Cold-Start problem on your Recsys에 대한 프로젝트이다.

주 수행 내용에는 Public 데이터셋을 활용해 입/출력을 정의하여 그에 따른 추천 시스템을 개발하고 개발된 상황에서 발생할 수 있는 콜드스타트 문제를 직접 정의 및 해당 상황을 시뮬레이션 할 수 있는 테스트 데이터셋을 구축하는 것이다. 그리고 해당 문제를 해결할 수 있다고 생각하는 방법을 기반으로 가설을 수립하여 추천할 baseline과 고도화된 알고리즘을 구성하고 검증하는 것이 목적이다.

1-2. 팀원 소개 및 프로젝트 수행 절차

표 1. 팀원 소개 및 역할

곽정무	PM, Negative Sampling Solution 고안
박준하	Cold User Retraining Solution 고안, 모델 Research 및 실험
박태지	데이터 수집 및 EDA, 모델 Research 및 실험
배현우	Baseline 구성 및 깃헙 코드 관리, 실험환경 고안, 데이터 전처리
신경호	Cold-Start problem 관측 및 정의, Side Information 및 Negative Sampling 실험
이효준	Cold, Warm 유저군별 EDA, Data Sampling, 추천 결과 후 분석



그림 1. 프로젝트 수행 절차

2. 데이터셋 개요

데이터셋 선정기준으로는 '활용할 보조 정보가 많은가?', '현실적으로 수집 가능한가?', '참고할 레퍼런스가 많은가?', '정제/가공되지 않은 데이터인가?', '조건에 제시된 유저와 아이템 수를 만족하는가?' 이 5가지를 선정하였고 그에 적합한 MovieLens 20M 데이터를 선정했다.

2-1. EDA

유저가 관심있는 아이템을 활용해 시청할 아이템을 추천하고자 평점 3 점 이상의 데이터만을 필터링한 뒤 EDA 를 진행했다.

EDA 결과 몇몇의 아이템에 유저의 평점이 밀집되어 있는 Long Tail 현상이 발생하는 것을 확인할 수 있었다. 또한 평가연도별 장르의 분포를 확인해봤을 때, 장르의 선호 경향은 1995 년을 제외하고는 동일한 분

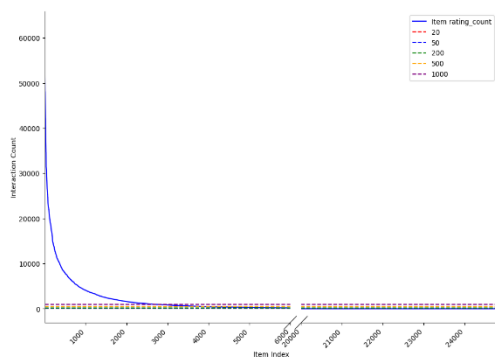


그림 2. 데이터셋의 Long Tail 현상

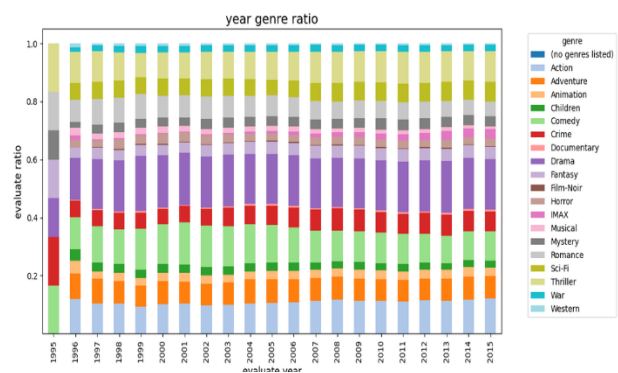


그림 3. 평가연도별 장르 분포

포를
지니
는
것도
확인
할
수
있었
다.

마지막으로 데이터의 Sparsity 의 경우 99.52%로 추천시스템의 데이터 중에서도 상당히 sparse 한 모습을 보였다.

3. Base RecSys

3-1. 추천 방법 및 평가 지표

추천 방법의 경우 일반적으로 사용되는 상황인 Top-10 추천을 선정했으며 평가 지표로는 Recall@10, nDCG@10 을 사용했다.

Recall@10 을 활용해 추천된 아이템 중 실제로 관심을 보인 아이템의 비율, nDCG@10 을 활용해 추천 순위에 가중치를 부여해 평가를 진행했다.

3-2. Base RecSys 선정이유

Base RecSys 로는 General Recommendation 을 선정했다.

실험은 유저별 상호작용을 8:1:1 로 랜덤 스플릿하는 환경에서 진행되었으며 이를 통해 빠르게 모델별 성능과 Cold Start 문제를 관측하고자 했다.

그리고 보조 정보가 존재하는 상황에서 이를 활용할 가치가 있는지를 판단해보고자 추가적인 분석을 진행했다. EDA 결과, 기본적으로 MovieLens 20M 데이터가 가지고 있는 변수들과 IMDB, TMDB 에서 추출한 변수들 모두 이력 수와 관련해 유의미한 특징을 보여주지 못했기에 추천 성능의 향상에 큰 도움이 되지 못할 것이라고 판단하여 Content based 추천을 Base RecSys 에서 제외했다.

또한 최근 추천시스템에서 주로 사용되는 Base RecSys 중 하나인 Sequential 추천 역시 고려대상 중 하나였으나 앞서 선정한 데이터셋에서는 활용하기 어렵다고 판단하여 추가로 제외했다.

그 근거는 MovieLens 20M 데이터의 경우 유저의 평가시점인 timestamp 변수를 가지고 있으나 해당 변수의 경우 시청 시점이나 시청 기간 등이 아니라는 점이었다. 그리고 EDA 를 진행한 결과 대부분의 유저가 동일한 날 여러 영화를 연달아 평가했음을 확인할 수 있었고 이를 바탕으로 해당 변수의 시퀀스적 특징을 신뢰하기 어렵다고 판단하였다.

4. Cold Start

4-1. Cold Start 설정 배경

해당 데이터셋에서는 Cold User와 Cold Item 두가지 문제를 동시에 확인할 수 있었다. 다만 앞서 언급했듯이 현재 데이터상에서는 EDA 과정에서 각 Side Information들에서 유의미한 정보를 얻을 수 없었으며 현업에서의 상황을 고려했을 때에는, Cold Item의 경우 다양한 Metadata를 보유하였기에 그에 대한 대응이 보다 원활하다고 느껴 연구 대상에서 제외하였다.

반면 Cold User의 경우 현업에서의 상황을 고려했을 때, 개인정보보호에 대한 문제로 인해 Metadata의 사용에 제약이 있어 그에 대한 대응이 보다 어렵다고 생각하여 본 프로젝트의 주 연구 대상으로 선정했다.

4-2. Cold Start User 확인

Cold Start 문제의 대상을 설정한 후 그에 대한 명확한 기준을 선정하기 위한 분석을 진행했다. 해당 데이터셋에서 Cold User 기준을 명확하게 선정하기 위해

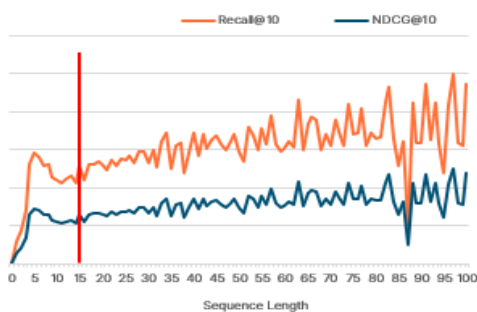


그림 4. LightGCN Sequence별 성능

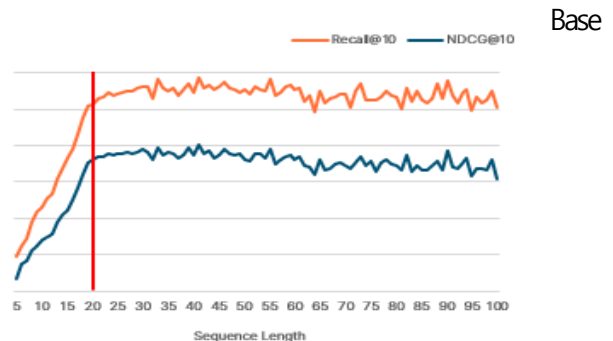


그림 5. SASRec Sequence별 성능

RecSys 뿐만 아니라 여러 모델을 활용해 분석한 결과는 아래와 같았으며 해당 결과와 앞선 EDA 결과를 바탕으로 본 프로젝트에서는 아이템과의 상호작용이 20 이하인 유저를 Cold User로 선정했다.

4-3. Cold Start User 특성

Cold Start User를 선정한 이후 Cold User와 Warm User 두 유저군에 대한 비교 분석을 진행했다. Cold User의 경우 Warm 유저에 비해 인기영화에 대한 시청 경향성이 높았음을 확인할 수 있었으며 최신영화에 대한 경향성의 경우 Cold User와 Warm User가 유사한 경향성을 가지는 것을 확인할 수 있었다.

표 2. 유저군별 인기영화 경향성

	Cold	Warm
Min	1	1
25%	151	3
50%	490	17
75%	1702	177
Max	62655	62655

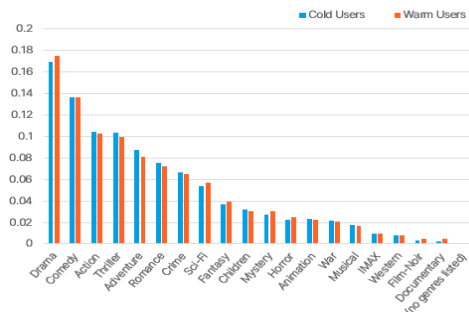


그림 6. 유저군별 장르 경향성

추가로 Cold User의 이력과 Warm User의 초기 이력을 비교했을 때, Warm 유저 전체 이력에 비해 초기 이력의 경우 인기 영화 경향성이 대폭 상승했으며, 보다

Cold User 시청 이력과 유사한 경향을 나타내는 것을 볼 수 있었다. 그리고 Cold User 시청 이력과 Warm User 초기 이력에서 공통으로 보이는 아이템에 대한 분석을 진행했을 때 가장 강한 인기영화 경향성을 나타내는 것을 확인할 수 있었다.

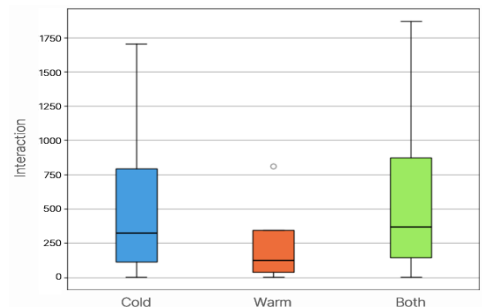


그림 7. Cold User와 Warm User 초기 시청 Item Interaction 통계량

5. Solution

5-1. Cold Start User 실험 환경 설정

Cold User에 대해 유저별로 8:1:1 Random Split을 진행할 경우, 해당 프로젝트에서 설정한 Cold User의 상호작용이 최대 20개이므로 Cold User의 테스트 샘플은 1~2개에 불과하다. 이러한 경우 Epoch마다 성능이 불안정해지며 테스트 샘플이 적음에 따라 평가 지표에 대한 신뢰성이 저하될 수 있다. 물론 교차 검증을 통해 이를 개선할 수 있으나 본 프로젝트에서는 보다 근본적인 문제라고 판단하여 실험 환경의 변경을 고안했다.

기존의 평점 3 점 이상인 데이터만을 사용한 평점기반 필터링 이후, 상호작용이 20 개 이상인 유저만을 사용하는 20 core 필터링을 적용하고 이어 상호작용이 아닌 유저를 기준으로 Train 8 : Valid 1 : Test 1 의 비율로 User split 을 진행했다.

그 후 Train 유저군의 모든 상호작용과 Valid, Test 유저군의 시간순 첫 K개(1~15 개) 상호작용을 Train 데이터로 활용하고 Valid, Test, 유저군의 K개 이후 이어지는 N개(N=5)의 상호작용을 Ground Truth 로 선정해 유사 Cold 유저로 변경했다.

이를 통해 시스템에 유입된 Cold User 의 상황을 잘 모사하여 평가 가능하며 유저별 시간 순서를 보전하여 현실 시나리오에 가까운 실험이 가능하다. 또한 Cold User 의 정답 수가 일정하게 설정되어 평가 신뢰도가 향상된다.

5-2. 개선 방안

1) Hard Negative Sampling

EDA 분석 결과 Cold User 들은 Warm User 들의 소비 이력과 비교해 가장 큰 차이로는 인기 있는 아이템을 주로 소비했다. Baseline 모델로 선정된 NCF 와 LightGCN 은 Top K Recommendation 으로 BPR Loss Function 을 사용했는데 BPR Loss Function 은 네거티브 샘플을 필요로 한다. Rendle(2009)에 따르면 Positive-Negative Sample Pair 를 구성할 때 Negative Item 으로는 아직 관측되지 않은, 즉 유저가 아직 소비하지 않았으며, 유저가 정말로 아이템에 관심이 없거나 아직 Missing Value 로서 미래에 소비될 아이템들로 구성된다.

Negative Sample 을 뽑는 방법으로는 보통 유저가 소비하지 않은 아이템에서 Random Sampling 하는 방식의 Uniform Random Sampling 이 자주 사용된다. 하지만, Cold User 의 이력 기반으로 확인했을 때 Cold User 가 소비하는 아이템들의 상위 75% 이력 횟수는 전체 User 가 소비한 아이템들의 상위 25%와 동일하므로 Cold User 에게 Uniform Random Sampling 은 전체 유저와 비교했을 때 Easy Negative Sampling 으로 볼 수 있어 다른 Negative Sampling 방식을 고려하였다.

인기 있는 아이템들로 후보군을 정해서 Random Sampling 을 진행하는 Popularity 기반 Negative Sampling 이 인기 있는 아이템을 주로 소비하는 Cold User 에게 Hard Negative Sampling 이라 판단하고 Train Data 에서 이력 횟수 상위 30% 아이템들에서만 Random Negative Sampling 을 진행했을 때 초기 성능은 Popularity 기반 Negative Sampling 이 좋았지만 NDCG 기준 Best Score 는 Uniform Random Sampling 이 더 좋았다.

Popularity 기반 Negative Sampling 이 초기 성능은 좋았지만 Uniform Random Sampling 의 고점을 넘지 못한 가설로 Popularity 기반 Negative Sampling 이 진행되면 진행될수록 Cold User 가 실제 소비할 만한 아이템을

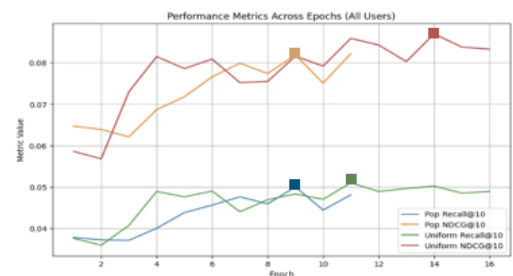


그림 8. Uniform vs 인기도 기반
Negative Sampling 비교

계속하여 Negative Sample 로 주기에 모델로서는 Popular 한 아이템들을 Cold User 들이 선호하지 않는다는 방향으로 해석할 수 있어 Uniform Random Sampling 으로 학습을 마친 이후 추가적으로 Popularity 기반 Negative Sampling 을 이용해 Hard Negative Sample 들을 학습했을 때 성능 향상이 일어날 수 있다는 가설을 세웠다.

표 3. Hard Negative Sampling 성능 변화

유저군	Recall@10	NDCG@10
1 Shot	0.0581(+23.6%)	0.0912(+5.8%)
3 Shot	0.0488(+5.9%)	0.0793(+2.1%)
5 Shot	0.0481(+3.2%)	0.0743(-3.3%)
10 Shot	0.0476(-6.1%)	0.0776(-13.0%)
15 Shot	0.0553(-1.8%)	0.0930(-5.9%)

위의 실험 설정 결과 Tradeoff가 발생함을 알 수 있었다. NDCG 는 올랐지만 Recall 은 내려가는 걸 확인할 수 있었고 1,3,5 Shot 유저군에 대해서는 성능이 오르는 모습을 보였으나 10,15 Shot 유저군에 대해서는 성능이 하락하는

Tradeoff를 관측할 수 있었다. 이러한 Tradeoff가 발생하는 원인으로는 User 의 이력이 더 초기일수록 더욱 더 인기 있는 아이템을 소비하는 성향이 있어 학습할 수 있는 이력이 더 적은 User 들에게 고안해낸 방식이 Hard Negative Sampling 이었다고 추측된다.

2) Cold User Embedding Retraining

LightGCN 을 전체 유저의 성능이 최대가 될 때까지 학습시킨 결과, 학습 후반부로 갈수록 cold user 의 성능이 급격하게 하락함을 확인할 수 있었다. 이는 cold user 에 대해 과도한 개인화가 이루어졌거나, 모델이 warm user 의 임베딩에 치우쳐져 학습하였기 때문이라고 추측한다.

물론 성능 감소가 일어나기 전에 학습을 멈추고 추론을 진행해도 되겠지만, 이렇게 얻은 cold user 의 임베딩 또한 최적이지 않음을 주장한다. Cold user 의 임베딩 입장에서 다른 warm user 와 item 의 임베딩은 일종의 학습 환경이다. 하지만 cold user 의 성능이 최대가 될 때 다른 유저와 아이템들이 충분히 학습되지 않았기에 cold user 가 좋은 학습을 할 수 없다는 것이다.

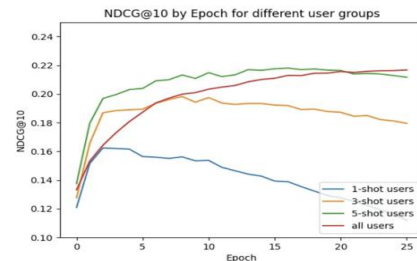


그림 9. Full Train 시킨 LightGCN 성능 그래프

표 4. 기존 성능과 Embedding Retraining 성능 비교

이러한 현상을 개선하고자, 모델이 완전히 학습을 완료한 뒤에 cold user 의 임베딩만을 정규분포 상에서 랜덤하게 초기화해

NDCG@10	Original (Best)	Original (Last)	Retrained (1 epoch)
1-Shot	0.1623	0.1119	0.1664
3-Shot	0.1981	0.1795	0.2092
5-Shot	0.2180	0.2116	0.2257
All	0.2167	0.2167	0.2217

서 재학습을 진행하는 방법을 시도하였다. 그 결과, 아래 그래프에서 관측할 수 있듯이 적은 epoch 만으로도 cold user의 성능이 학습 완료 시점에 비해 급격히 개선되었다.

눈여겨볼 점은 재학습을 진행했을 시 도달하는 성능이 이전 모든 epoch에서의 cold user 성능보다도 높았다는 것이다. 이러한 관측은 다른 임베딩들이 최적일 때 cold user의 학습이 개선될 것이라는 가설을 뒷받침해준다.

6. 회고

추천 시스템에서 보편적으로 Cold-Start 문제는 존재한다. 실험을 진행하기에 앞서 다양한 Base RecSys를 통하여 우리가 사용하는 MovieLens 20M 데이터에도 관측이 된다는 것을 확인한 후 솔루션을 탐구하였다. 평점 3점을 threshold로 선정하여 필터링을 진행하여 implicit한 데이터로 변환하였다.

top@k metric이 콜드 유저군과 웜 유저군 사이의 평가 형평성에 문제가 있다는 것을 확인하여 실험 환경을 다시 구축하고 다양한 방법을 시도한 끝에 negative sampling과 user embedding retraining 두 방법을 고안하였다. 해당 솔루션들은 긍정적인 결과를 내었지만 대표적으로 3가지의 문제점을 지니며 이에 대한 후속 연구 방안을 제시한다.

첫째, N shot 상황에서의 보편화가 필요하다. 현재 실험에서 few shot에 대한 상황에서는 성능이 증가하지만 나머지 상황에서는 성능이 하락하는 trade-off가 관측된다. 해당 모델만을 사용하면 전체적인 성능이 하락하기 때문에 Cold-start 상황이 아닌 경우에도 일관적인 성능 향상이 필요하다.

둘째, 인기도 기반 item이 성능이 좋다. cold-user의 특성 상 인기도가 높은 item이 성능이 상승하는 경향을 볼 수 있다. 인기도 기반으로만 추천을 진행하면 같은 아이템만 지속적으로 노출이 될 수 있으므로 다양화된 추천이 필요하다.

셋째, Valid Set의 크기에 따라서 Solution의 유효성이 변화한다. user embedding을 재 학습하는 방법론에서 Valid Set의 크기에 따라서 성능 개선 및 방향성의 결과가 변화한다. 그러므로 해당 Solution의 유효성을 보장하기 위하여 다방면의 추가적인 실험이 필요하다.

이번 프로젝트를 진행하며 문제 정의 및 해결 방안 고안, 실험 등 전반적인 프로세스를 경험하였다.

References

Understanding Negative Sampling in Graph Representation Learning

Zhen Yang, Ming Ding, Chang Zhou, Hongxia Yang, Jingren Zhou, Jie Tang. 2020

MeLU: Meta-Learned User Preference Estimator for Cold-Start Recommendation

Hoyeop Lee, Jinbae Im, Seongwon Jang, Hyunsouk Cho, Sehee Chung. 2019

Generalizing from a Few Examples: A Survey on Few-Shot Learning

Yaqing Wang, Quanming Yao, James T. Kwok, and Lionel M. Ni. 2020.

Fast Adaptation for Cold-Start Collaborative Filtering with Meta-Learning

T. Wei et al. 2020.

Cold-Start Recommendation towards the Era of Large Language Models (LLMs): A Comprehensive Survey and Roadmap

Weizhi Zhang, Yuanchen Bei, Liangwei Yang et al. 2025.

개인회고 (곽정무_T7504)

기존의 프로젝트에서 깔끔하게 정제된 데이터셋이 주어지고 향상시켜야 되는 성능 지표, 그리고 성능을 측정할 실험 환경들이 미리 조성되어 있던 반면, 이번 프로젝트는 모든 절차에 있어 논리적인 흐름에 맞게 진행이 되고 있는지 매번 의문을 품어야했고 이러한 부분에서 프로젝트를 진행하며 가장 어려움을 겪었다.

이번 프로젝트에서도 주로 맡은 역할은 리서처였다. Cold Start 문제에 대해 학계에서 어떤 진전이 있는지 전체적인 맥락을 파악하기 위해 Cold Start Survey 논문을 읽었고 다루고자 하는 문제의 범위를 Cold User 로 제한하기 위해 Few Shot Learning Survey 도 읽고 메타 러닝 방법론을 적용하고자 관련 논문을 여러 편으로 읽었으나 하드웨어적인 한계로 방법론을 적용해 보지 못한 것은 더 효율적인 엔지니어링을 통해 한계를 극복해 봄 직했으나 역량의 부족함으로 인해 아쉬움이 많이 남았다.

메타 러닝의 Parameter Initialization 에 대한 아이디어들은 Warm User 에 대해 모델을 사전학습하고, Warm User 들의 풍부한 데이터로 User Embedding 이 잘 학습되었다는 가정하에 Cold User 들의 이력으로 Fine Tuning 을 진행할 때 0-shot 상황에서는 평균적인 유저가 볼 법한 아이템을 추천하는 것이 맞다는 이론 하에 Warm User 의 Embedding 평균으로 학습하는 아이디어로 이어졌는데 비록 결과가 생각만큼 나오진 않았지만 데이터의 성질과 도메인 지식을 활용해 논리적인 아이디어로 이어져 과정은 만족스러웠다.

또한, Popularity 기반 Negative Sampling 솔루션을 도출해 내는 과정에서 Negative Sampling 을 연구하던 팀원이 발견한 사실 및 가설들을 취합해 Cold User 들에게 Hard Negative Sampling 이 효과적일 수 있다라는 가설로 이어졌고 잦은 회의를 통해 팀원들이 함께 아이디어를 발전시켜 나가 최종 솔루션으로 이어진 흐름은 팀으로 프로젝트를 같이 진행하면서 집단의 경험과 지식을 활용해 한단계 발전하는 모습을 목도할 수 있어서 좋았다.

프로젝트를 진행하면서 가장 아쉬웠던 부분은 프로젝트에서 항상 PM 역할을 도맡아 프로젝트 진행을 조율했는데 이번 프로젝트에서는 프로젝트의 진행 방향을 대략적으로라도 파악하기가 어려웠고 개인의 일처리에 바빠 팀의 프로젝트 방향을 조율하는데 실패하여 앞으로 개선할 부분이라고 생각된다. 특히나, 문제를 정의하고 문제의 범위를 한정시키고 난 뒤에는 프로젝트 구성에 대한 청사진이 머리 속에 그려진 만큼 PM 으로서 문제에 대한

핵심을 빠르게 파악하고 솔루션을 적용할 수 있는 범위를 고려해 팀을 이끄는 게 필요하다고
생각된다.

실무에 가까운 프로젝트를 진행하면서 처음이었기에 많은 고민과 느린 진행에 따른
실망감도 따랐지만 일찍이 경험할 수 있어 귀중한 경험이었다고 생각된다. 팀원들과
데드라인 마지막 날까지 힘을 합쳐 끝까지 몰입하며 나름의 솔루션을 도출해 낼 수 있던
부분은 잊지 못할 경험이었다.

개인회고 (박준하_T7523)

이번 프로젝트는 많은 실험과 연구를 통해 크게 성장할 수 있었던 값진 경험이었다. 여러 시도 중 가장 인상 깊었던 것은 관련 논문을 읽고 negative sample의 분포를 다방면으로 변형시켜본 경험으로, 특히 easy negative sampling에 집중하였었다. 비록 데이터셋에서의 일관성 문제와 주장한 이론의 오류로 실패하였지만 처음으로 체계적인 연구와 실험을 진행해본 경험이라 앞으로 연구자의 길을 걷는데 중요한 초석을 쌓았다고 믿어 의심치 않는다.

Negative sampling에 많은 시간을 투자했지만 마지막에 짧은 시간 동안 진행한 finetuning, 혹은 retraining 방법론이 역설적이게도 팀에 기여할 수 있었던 가장 효과적인 접근이었다. 학습이 종료된 뒤에 cold user의 임베딩을 초기화 후 다시 학습한다는, 상대적으로 간단한 방법임에도 큰 효과를 보여 놀라웠다. 짧은 시간을 투자했지만 해당 실험이 성공한 이유를 지금와서 곰곰히 생각해보니 실험의 가설이 탄탄하고 직관적으로 잘 뒷받침되었다는 점이 큰 차이를 만들어냈던 것 같다. 이전 negative sampling 실험들은 가설 자체도 모호했고 수식을 장황하게 전개했지만 직관적인 해석이 부족했다. 부스트캠프를 마치고도 앞으로도 다양한 실험을 진행하게 될텐데 이러한 깨달음을 기억하고 가설을 세울 때 명심하자고 다짐한다.

프로젝트를 진행하면서 느꼈던 가장 큰 어려움은 데이터셋의 크기와 GPU 자원의 한계였다. 이전까지는 user-item interaction 그래프를 단순히 밀집행렬로 GPU에 올려도 별 문제가 되지 않았는데, 대용량 데이터인 MovieLens 20M 데이터로 실험하려니 당연하게도 VRAM 용량이 부족했다. 실험 시간도 문제였다. 1M 데이터의 경우에는 빠르면 10분 안에 실험을 완료할 수 있었는데 20M 데이터의 경우에는 A100으로도 최소 5~6시간 정도 걸렸다는 점이 발목을 크게 붙잡았다. 당연히 현업 상의 데이터는 이보다도 더 클테니 distributed learning과 같은 방법론들이 괜히 중요한 것이 아님을 실감하였다.

비록 많은 어려움과 실패를 겪었지만, 이번 프로젝트에서 얻은 값진 경험과 깨달음은 앞으로 좋은 연구를 이어나갈 든든한 밑거름이 될 것이라 확신한다.

개인회고 (박태지_T7524)

이번 프로젝트의 주된 목표는 cold-user에 대한 추천을 개선하는 것이었다. 프로젝트는 데이터셋 구성부터 시작하여 모든 과정을 스스로 결정해야 했기 때문에, 기존의 정형화된 방법론을 적용하기 어려웠다. 이로 인해 여러 번의 시행착오가 발생하였으며, 데이터의 특성 및 품질에 따라 모델 성능이 크게 좌우됨을 직접 체감하였다.

모델 구현 단계에서 환경에 맞게 모델을 적용하는 데 한계가 있었다. 먼저, 빠른 실험을 위해 MovieLens 1M 규모의 데이터를 사용하였고, 이후 20M으로 변경하자 일부 모델은 예상보다 오랜 실행 시간을 보였기에 실제 적용에 어려움이 발생하였다. 이에 따라 모델의 효율성을 높이기 위한 여러 수정 작업을 진행하였다. 또한 실험 환경을 처음부터 동일하게 구성하려는 노력에도 불구하고, 데이터와 상황에 맞는 최적의 환경을 찾는 것이 어려웠다. 이로 인해 초기 실험 결과와 후속 결과 간의 비교 분석에 어려움이 발생하였다.

Multi-VAE를 기반으로 하여 cold start 문제를 해결하고자 하였다. cold user가 인기 아이템에 쏠리는 현상을 막기 위해 기존의 모델 Loss에 Diversity Loss를 도입하였다. 사용자의 latent 표현 벡터의 분산을 최대화하는 방식으로, 다양한 추천을 유도하고자 하였다. 하지만 후분석 결과 cold user에게 추천되는 아이템이 인기 있는 아이템에 쏠리는 현상을 근본적으로 막을 수 없었고, 해당 방식은 사용할 수 없었다.

이번 프로젝트는 현업에서 진행되는 방식과 유사한 상황을 경험할 수 있었으며, 데이터셋 구성부터 모델 구현, 후분석까지 전 과정을 직접 관리하는 어려움을 체감하였다. 여러 시행착오와 수정 과정을 통해 기존과는 다른 경험을 할 수 있어 좋은 바탕이 될 것이라고 생각한다.

개인회고 (배현우_T7525)

현실의 문제는 생각보다도 더욱 어려웠다. 데이터셋을 선정하고, 베이스 렉시스를 구성하고, 쿨드 스타트를 관측 및 정의하고 개선을 시도하기 전까지의 과정이 예상보다도 훨씬, 훨씬 오래 걸렸다. 캐글의 컴페티션 같은 경우엔 데이터, 문제, 메트릭 등 모든 것이 이미 정해져 있어 바로 실험부터 진행을 하기만 하면 되었다. 그러나 모든 것을 직접 하나하나 정하며 진행하고, 각각을 정할 때 많은 고민과 근거들이 필요하게 되었다.

실험 환경에 대해서도 고민이 많을 수밖에 없었는데, 검증 방법도 각양각색이고, 입력의 형식, 네거티브 샘플링 등 어느 것에 맞춰야 할지 쉽게 정할 수가 없었다. 일단 실험은 진행해야 했기에 결국 각자 다른 환경에서 다른 모델을 실험해 보는 상황이 되었는데, 그로 인해 체계적인 실험이 불가능 했다. 공정하고 신뢰성이 있을 것이라고 생각되는 실험 환경을 고안하여 적용하려고 했으나 학습법이 달라 사용이 불가능한 경우가 생겼다. 그래도 이 모든 과정에서 실험 환경에 대한 고민을 많이 하고 배워 얻은 것이 많다고 느꼈다.

현업에서처럼 명확하지 않은 환경에서 프로젝트를 알아서 진행하려다보니 팀 전체적으로 방향을 하게 되었고, 그에 따라 진행 과정이 매끄럽지 않고 중구난방이었다. 그 와중에 멘토님에게 조언을 들으며 방향성을 정하고 나니 머리가 맑아지는 느낌이었는데, 전체적인 그림을 가지고 팀을 이끌 수 있는 사람의 필요성을 크게 느끼게 되었다. 능력의 부족으로 그런 조율을 하는 데에 기여하지 못한 점이 아쉬웠다.

결과적으로는 과정은 만족스러운 프로젝트였다. 솔루션을 찾기 시작하기 전까지의 과정도, 솔루션을 찾는 과정도 쉽지 않았지만 얻을 수 있던 성과를 정리하고, 후분석까지 진행한 것을 바탕으로 결과를 발표하였는데, 그 과정에서 모든 팀원이 열심히 힘을 합쳐 잘 마무리할 수 있었던 것 같다. 팀워크를 발휘하여 끝맺음을 잘 한 부분이 대단한 성과는 아니더라도 수상을 할 수 있던 원동력이 아닐까 한다.

개인회고 (신경호_T7530)

기존의 대회형식의 프로젝트는 데이터, 문제, 메트릭 등 주어진 환경에서 개선을 하는, 말그대로 방향을 알려주고 그 길을 따라서 진행하기만 하면 되었다. 그러나, 이번 프로젝트는 해결해야 하는 문제만 알려주고 나머지는 스스로 생각하고 고안해야 하는 문제였다. 문제 정의부터 해결방안, 평가 방법 등 하나부터 열까지 직접 생각하고 고민하는 프로젝트는 실무에서 문제를 해결하는 방식과 비슷하며 많은 생각과 성장을 이루었다.

프로젝트를 시작하면서 가장 중요하게 여긴 부분은 문제를 정의하는 것이었다. Cold-start 문제가 RecSys 에 존재한다는 것은 공공연한 사실이지만 실제로 관측하여 정의하는 것은 다른 영역의 문제이기 때문이다. 그래서 실험을 바탕으로 객관적인 지표를 도출하는 것을 중요하게 생각하였다. 이를 위해 Graph 기반인 LightGCN 과 Sequential 모델인 SASRec 을 통하여 해당 문제가 존재함을 증명하였다. 오픈 소스를 현재 실험 환경에 맞춰서 수정하는 과정이 힘들었지만, 이 과정을 통해 타인의 코드를 분석하여 현재의 환경에 맞출 수 있는 능력을 기를 수 있었다고 생각한다.

그 후에는 다양한 실험을 진행하였다. Content 를 이용한 솔루션을 위하여 Transformer4Rec 과 DIF-SR 에 대한 논문을 읽어보며 현재 상황에 적합한 모델을 선정하여 고안하였으나 구현과정에 어려움을 겪었고, 팀원과 회의 끝에 Sequential 모델은 제외하기로 하여 아쉬움이 남았다. 또한, LightGCN 을 개선한 LightGCNpp 모델도 실험을 진행하여 했으나 시간관계상 진행하지 못하였다는 아쉬움이 남았다.

아쉬움이 많은 만큼 값진 경험도 얻었다. 실제 현업에서 문제를 해결하는 프로세스와 유사한 경험을 한 만큼 어떤 관점을 가져야 하는지 배울 수 있었으며 단순히 코드를 구현하는 부분 외에 많은 것들을 고려하고 투자해야 한다는 것을 알 수 있었다.

개인회고 (이효준_T7545)

이번 프로젝트는 기업 해커톤이자 부스트캠프에서의 마지막 프로젝트로, 부스트캠프 기간 동안 설정한 데이터 분석가라는 커리어 방향성에 맞춰 업무를 수행하는 데 집중했다. 특히, 프로젝트 주제에 적합한 데이터를 선정하고, 주요 변수를 선별하여 분석한 후, 인사이트를 도출하고 시각화하는 역할을 담당했다.

프로젝트 초반에는 "Cold Start"라는 주제와 팀의 강점을 연결할 수 있는 데이터를 선정하는 것이 최우선 과제였다. 우리 팀의 강점은 부스트캠프 초반부터 지속적으로 진행해온 추천 시스템 관련 논문 리뷰였으며, 이를 활용하기 위해 많은 연구에서 사용된 MovieLens 데이터를 선정했다. 이후, 추가적인 외부 데이터를 확보하기 위해 IMDB와 TMDB에서 추가적인 변수를 추출하여 활용했다.

데이터 선정 후, 의미 있는 변수를 선별하기 위해 EDA를 진행했다. 먼저, IMDB와 TMDB에서 가져온 변수들의 특성을 분석한 후, 개봉 연도, 인기도 등의 요소가 해당 변수에 미치는 영향을 분석하고 시각화했다. 또한, 데이터의 수집 방식을 조사하여 배우, 감독 등 다중값을 가지는 변수들이 어떻게 정리되어 있는지 파악하고, 이를 효과적으로 활용할 수 있는 방안을 모색했다.

추가적으로, 추천 시스템의 Cold Start 문제를 해결하기 위해 Base RecSys 실험을 바탕으로 유저를 Cold User와 Warm User로 분류하고 각 그룹에 대한 추가적인 EDA를 진행했다. 분석 결과, Cold User는 Warm User에 비해 인기 있는 아이템을 선호하는 경향이 뚜렷하다는 점을 발견할 수 있었다. 이를 통해 주제와 관련성이 낮은 변수를 필터링하여 모델의 성능을 개선하는 데 기여했다.

이번 프로젝트에서 새로운 시도로 진행한 부분은 실험의 효율성을 높이기 위한 데이터 샘플링이었다. 원본 데이터는 약 20M에 달하는 방대한 규모였으며, 이로 인해 실험 시간이 지나치게 길어지는 문제가 발생했다. 이를 해결하기 위해 클러스터링 및 층화 추출기법을 활용하여 원본 데이터와 유사한 샘플 데이터를 생성했다. 이를 바탕으로 보다 신속하고 효율적인 실험을 진행할 수 있었다.

프로젝트 후반부에는 팀원들이 수행한 다양한 실험 결과를 후분석하고, 프로젝트의 전반적인 내용을 효과적으로 전달하기 위한 시각화 작업을 수행했다. 단순한 그래프나 표를 활용하는 데서 벗어나, 분석 결과를 보다 직관적으로 전달할 수 있는 방법을 고민하며 데이터 시각화의 효과를 극대화했다.

이번 프로젝트를 통해 가장 크게 배운 점은 실제 데이터 분석 프로젝트의 흐름을 경험할 수 있었다는 것이다. 프로젝트 초반에는 가설을 어떻게 설정하고, 이를 검증하며 프로젝트를 진행해야 하는지에 대한 고민이 많았다. 이러한 과정에서 기업 멘토님의 기술 블로그 글과 조언을 참고하며 프로젝트의 방향성을 설정했다. 이를 통해 현업에서 프로젝트를 진행하는 방식과 실험을 통해 정당성을 확보하는 과정을 직접 경험할 수 있었다.

또한, 시각화 기법에 대한 깊이 있는 고민을 할 수 있었다. 일반적인 그래프나 표를 사용하는 것을 넘어, 실험 결과를 어떻게 표현해야 상대방이 더 쉽게 이해하고 설득될 수 있을지를 고려하며 분석 결과와 인사이트를 효과적으로 전달하는 방법을 익혔다.

마지막으로, 문제를 정의하고 데이터를 활용하는 역량을 키울 수 있었다. 사전에 정의된 Metric이나 데이터가 없는 상태에서, 어떤 데이터를 사용할지, 변수들은 어떻게 활용할지, 그리고 발생하는 문제를 데이터 기반으로 해결하는 과정을 거치며 데이터 리터러시를 한층 더 향상시킬 수 있었다.

이번 프로젝트를 통해 데이터 분석가로서의 실무적인 경험을 쌓았으며, 분석 역량뿐만 아니라 실험 설계, 문제 해결, 그리고 데이터 기반의 의사결정 역량을 강화할 수 있었다. 앞으로도 이러한 경험을 바탕으로 스스로의 커리어를 더욱 발전시키는 것이 개인적인 목표이다.