

2023 인하 인공지능 챌린지

Team : MI-Lab

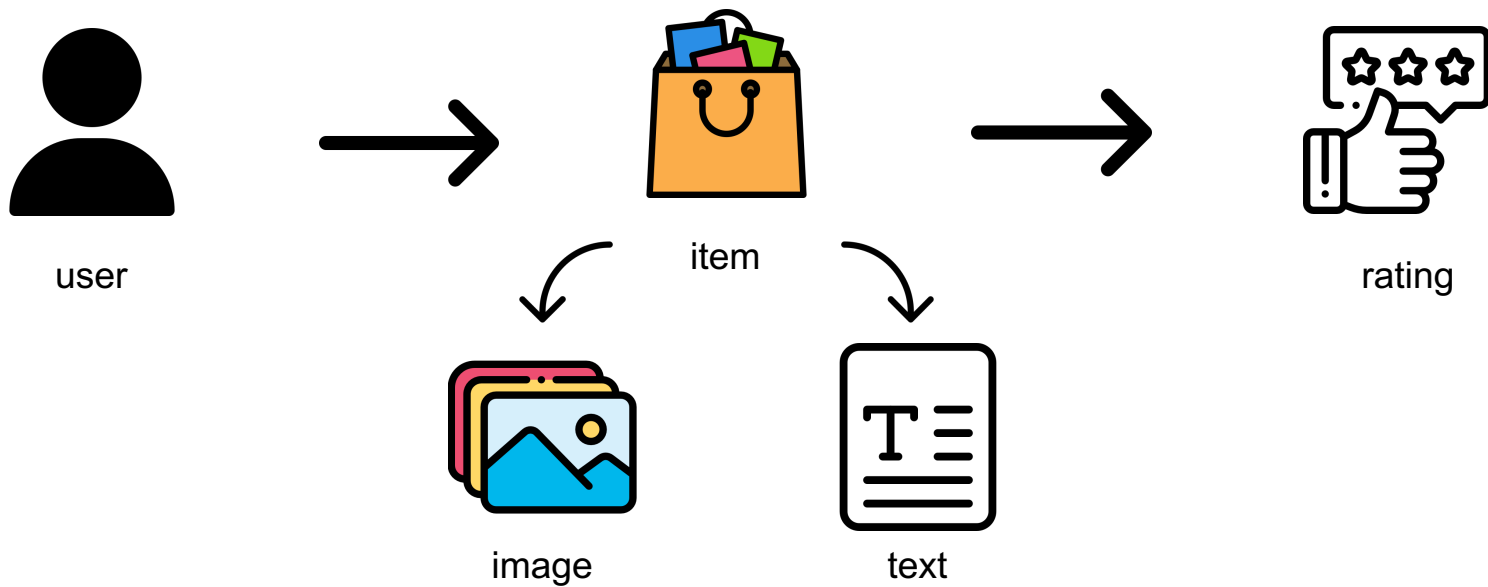
12181898 허건혁
12191870 정봉기
12191876 한승현



목차

1. 데이터
2. 모델
3. 결과
4. 실패한 시도

대회 설명



평가 산식

평가 산식 : **NDCG@50(Normalized Discounted Cumulative Gain)**

$$DCG_u = \sum_{i=1}^k \frac{relevance_i}{\log_2(i+1)}$$
$$IDCG_u = \sum_{i=1}^k \frac{relevance_i^{opt}}{\log_2(i+1)}$$
$$NDCG_u = \frac{DCG_u}{IDCG_u}$$

$relevance_i$ 는 3점 이상의 제품 선호도(Rating)를 이진화하여 사용되며, $relevance_i^{opt}$ 는 $relevance_i$ 를 선호도 순으로 정렬하여 사용합니다.

예를들어, A-E에 대해 u_1 의 $relevance_{u,1} = \{5, 3, 2, 4, 0\}$ 와 같을 때, $recommend_{u,1} = \{A, C\}$ 의 NDCG는 $(\frac{1}{\log_2(1+1)}) / (\frac{1}{\log_2(1+1)} + \frac{1}{\log_2(2+1)}) \cong 0.6131$ 으로 측정됩니다.

Public Score : 전체 테스트 데이터 중 user_id 별 item_id의 30% (대회 기간 중 실시간 점수 공개)

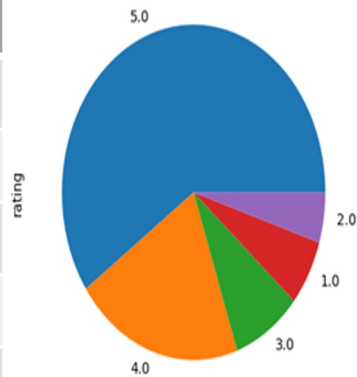
Private Score : 전체 테스트 데이터 중 user_id 별 item_id의 나머지 70% (대회 종료 후 점수 공개)

데이터

User (0~192403)	Value
Count	192403.00
Mean	6.52
Std	6.93
Min	3.00
Q1	3.00
Q2	5.00
Q3	7.00
Max	345.00

Item (0~63001)	Value
Count	62989
Mean	19.92
Std	56.16
Min	1.00
Q1	5.00
Q2	8.00
Q3	17.00
Max	3688.00

Rating	Counts	Ratio
5	749623	60%
4	258483	21%
3	105720	8%
2	60842	5%
1	79773	6%



모델 선정

Dataset	# Users	# Items	# Interactions	Sparsity
Baby	19,445	7,050	160,792	99.8827%
sports	35,598	18,357	296,337	99.9547%
FoodRec	61,668	21,874	1,654,456	99.8774%
Elec	192,403	63,001	1,689,188	99.9861%

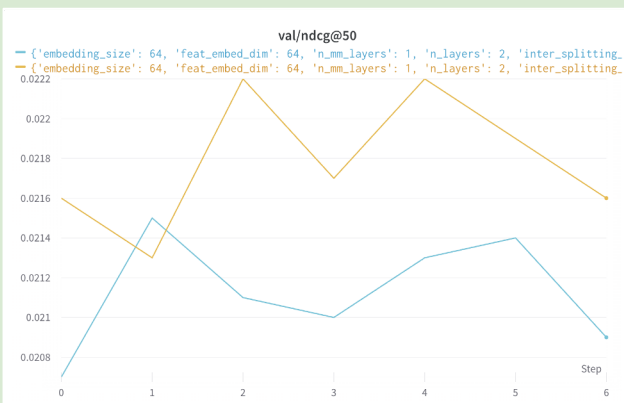
Inha	192,403	62,989	1,254,441	99.9896%
-------------	----------------	---------------	------------------	-----------------

모델 선정

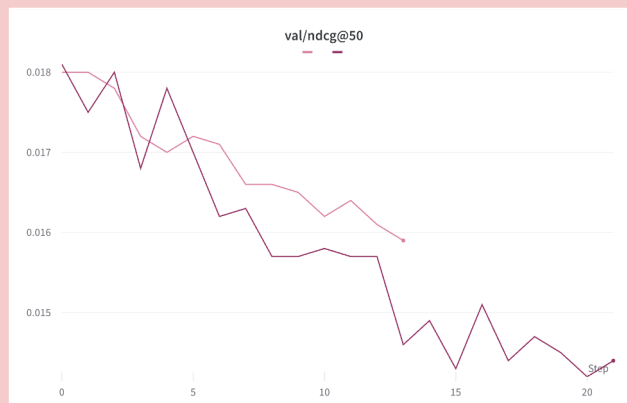
Dataset	Model	Recall@10	Recall@20	Recall@50	NDCG@10	NDCG@20	NDCG@50
Elec	BPR	0.0235	0.0367	0.0621	0.0127	0.0161	0.0212
	LightGCN	0.0363	0.0540	0.0879	0.0204	0.0250	0.0318
	VBPR	0.0293	0.0458	0.0778	0.0159	0.0202	0.0267
	MMGCN	0.0213	0.0343	0.0610	0.0112	0.0146	0.0200
	DualGNN	0.0365	0.0542	0.0875	0.0206	0.0252	0.0319
	GRCN	0.0389	0.0590	0.0970	0.0216	0.0268	0.0345
	LATTICE	-	-	-	-	-	-
	BM3	0.0437	0.0648	0.1021	0.0247	0.0302	0.0378
	SLMRec	<u>0.0443</u>	<u>0.0651</u>	<u>0.1038</u>	<u>0.0249</u>	<u>0.0303</u>	<u>0.0382</u>
	ADDVAE	0.0451	0.0665	0.1066	0.0253	0.0308	0.0390
	FREEDOM	0.0396	0.0601	0.0998	0.0220	0.0273	0.0353

다양한 모델 시도

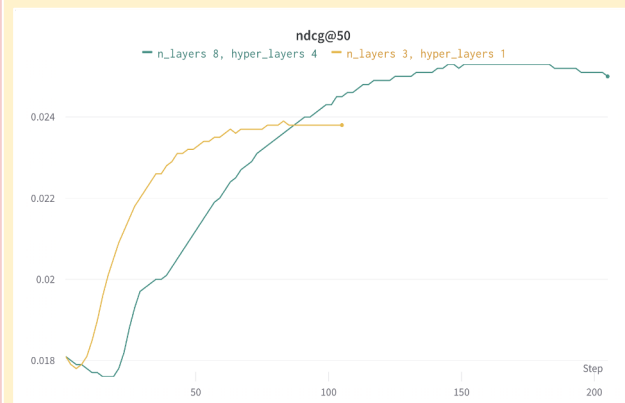
DRAGON



DualGNN



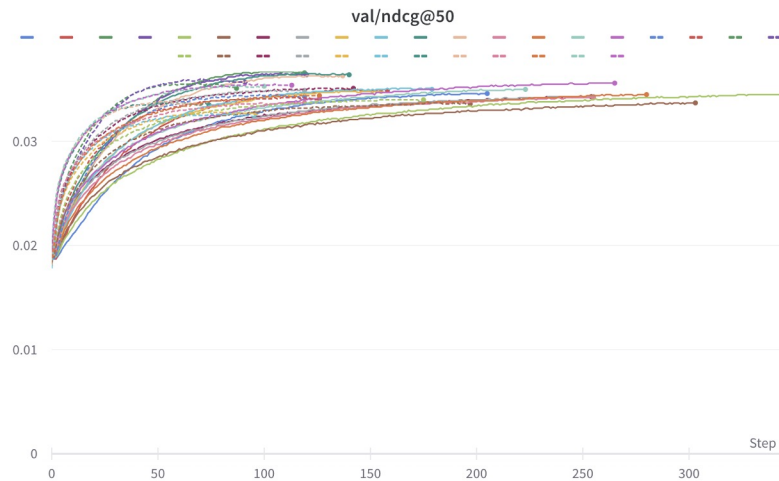
NCL



이외에 FM, FREEDOM, Multi-VAE, BSPM 등 좋은 성능을 보여주지 못함

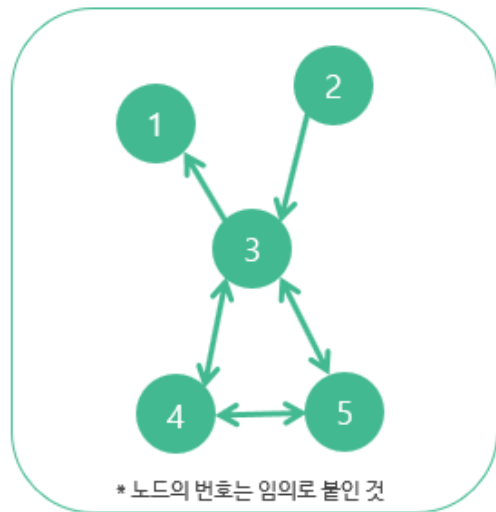
최종 선택 : BM3

Parameter	Value
Embedding_Size	64,128
Feat_embed_dim	64, 128
N_layer	1, 2, 3, 4
Drop_out	0.3, 0.5

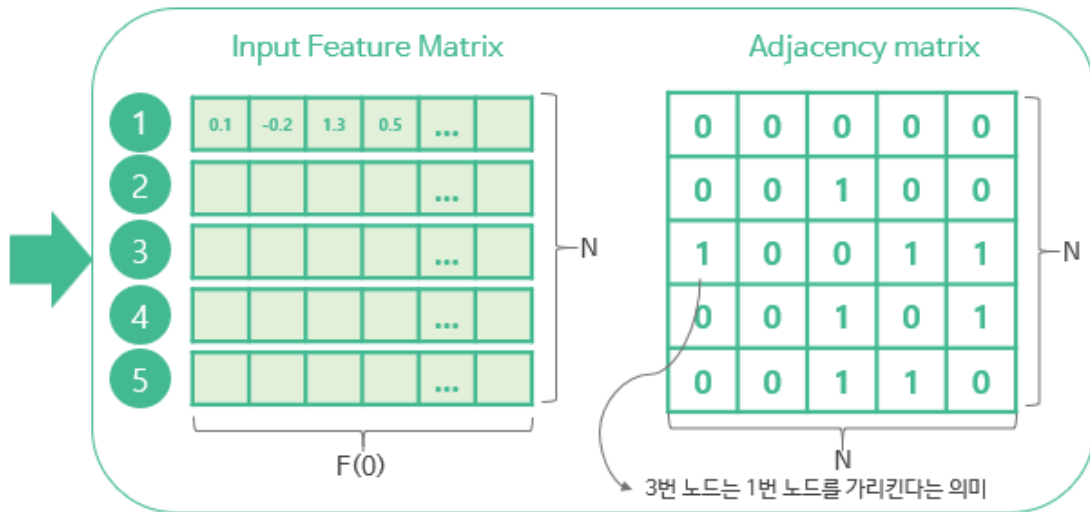


GCN

graph G



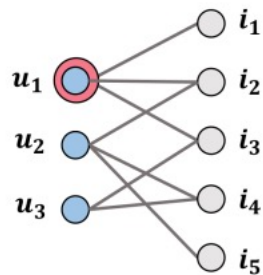
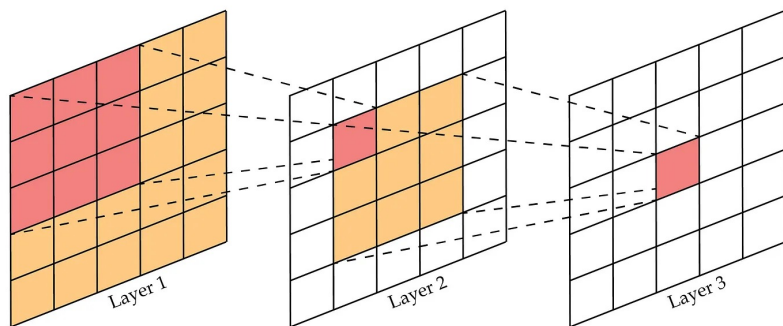
GCN INPUT



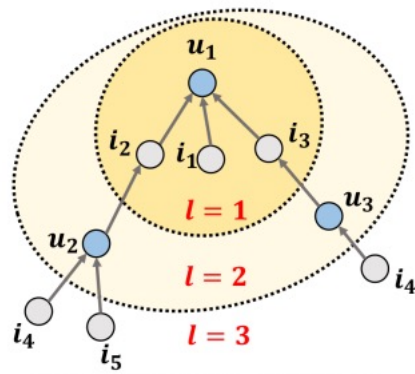
$$f(H^i, A) = \sigma(D^{-1}AH^iW^i)$$

GCN

Receptive Field in Convolutional Networks



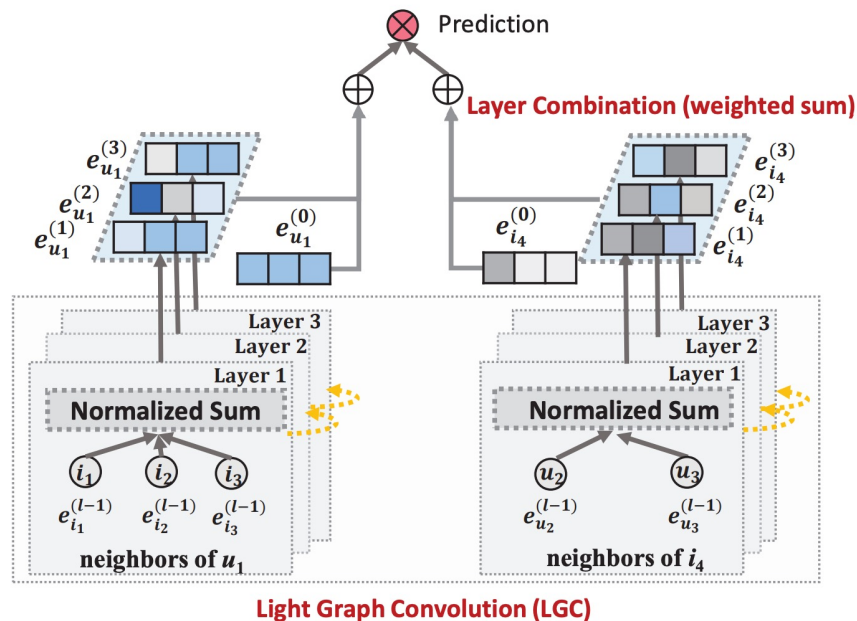
User-Item Interaction Graph



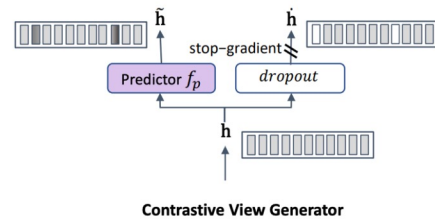
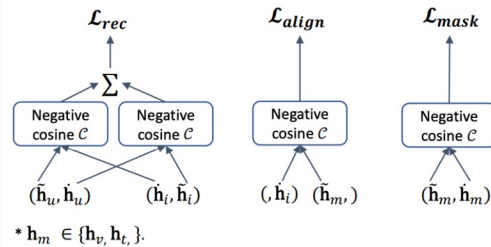
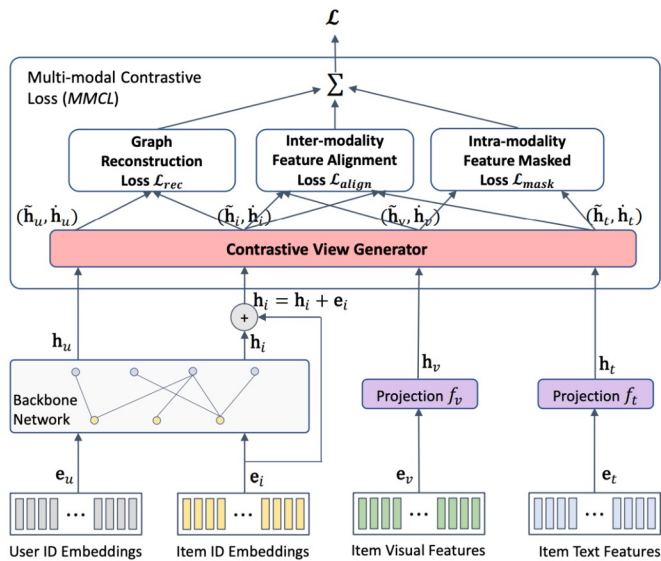
High-order Connectivity for u_1

Light GCN

$$\mathbf{E}^{(k+1)} = (\mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}) \mathbf{E}^{(k)}$$



BM3



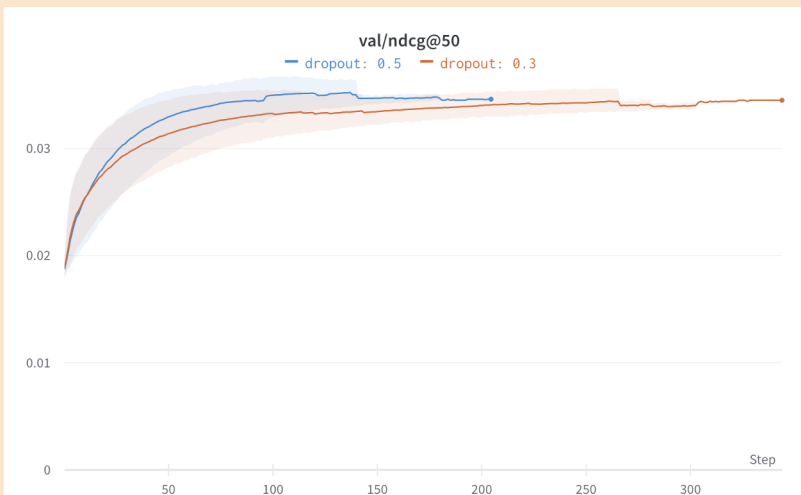


5-fold

user - item Interaction 3인 경우	
Train	Valid
A, B	C
A, C	B
B, C	A
위 3가지 경우에서 2개 추가 선택	

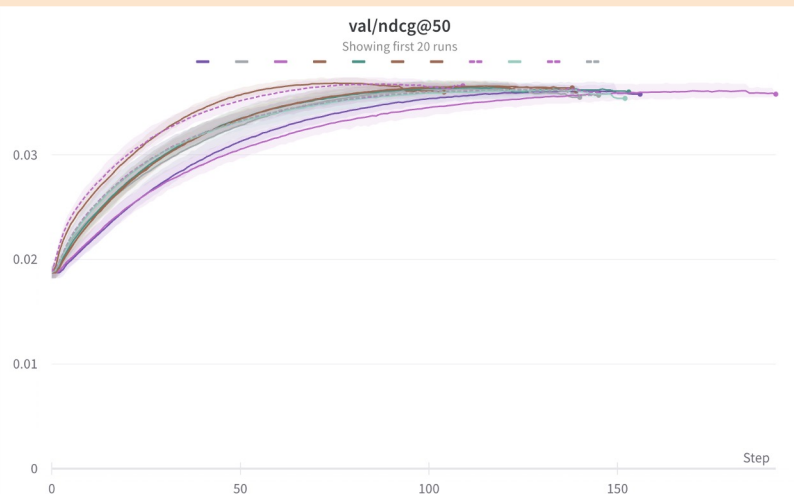
user - item Interaction 4인 경우	
Train	Valid
A, B, C	D
A, B, D	C
A, C, D	B
B, C, D	A
위 4가지 경우에서 1개 추가 선택	

Hyperparameter Tuning



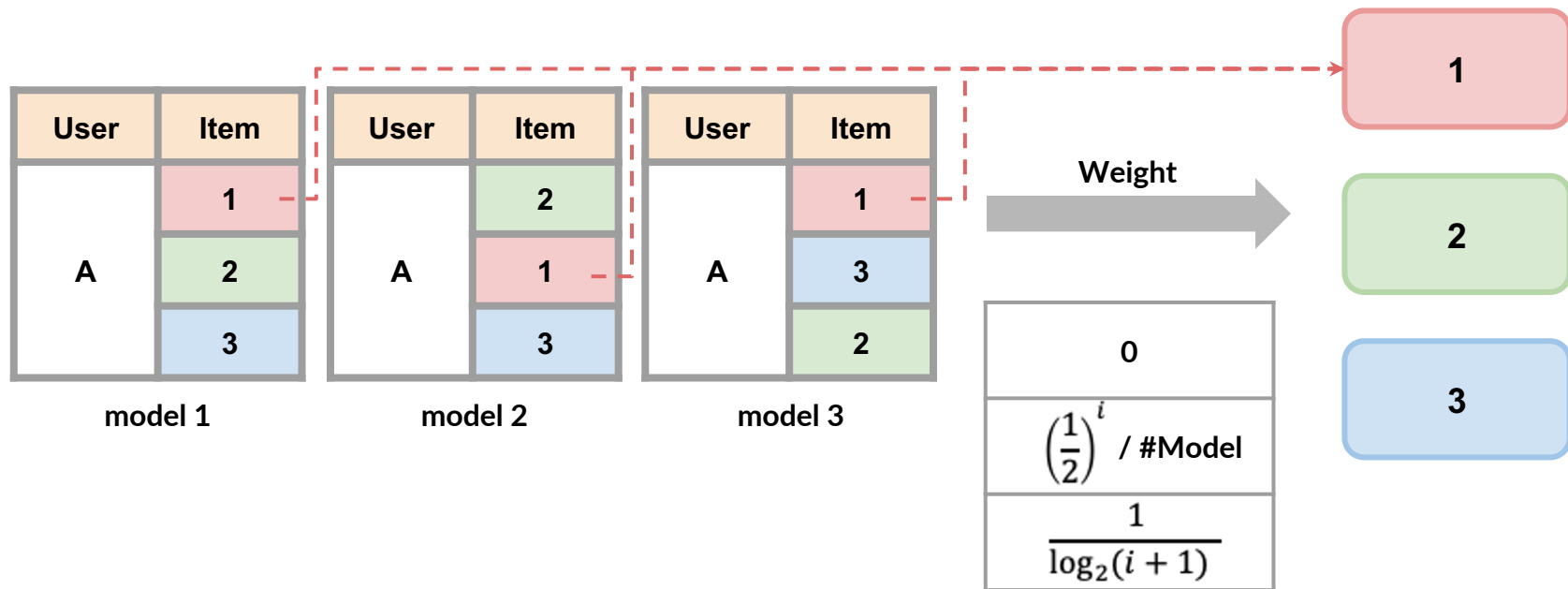
Parameter	Value
Embedding_Size	64,128
Feat_embed_dim	64, 128
N_layer	1, 2, 3, 4
Drop_out	0.3, 0.5

Hyperparameter Tuning

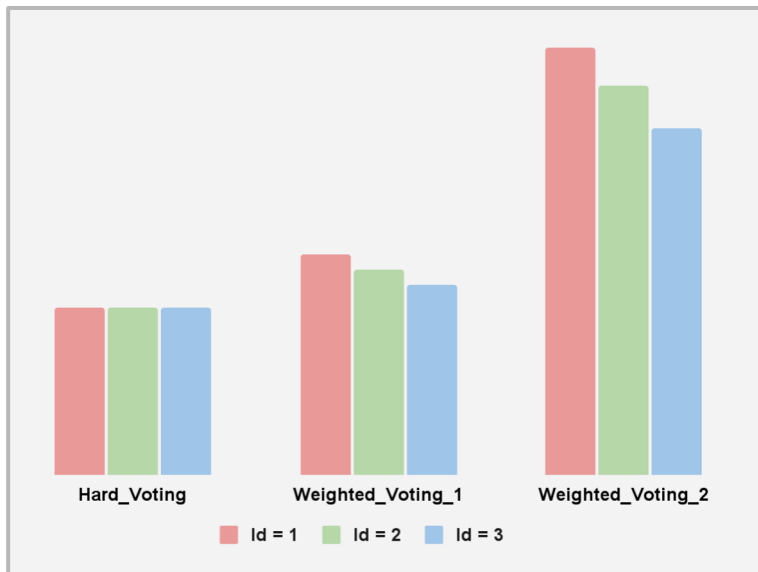


N_layer	Embedding	Feat_embed	ndcg@50
4	256	128	0.0369
3	256	128	0.0368
4	128	128	0.0366
4	128	256	0.0366
4	128	64	0.03649

결과



결과



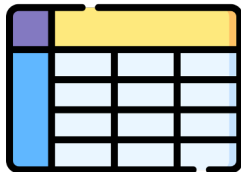
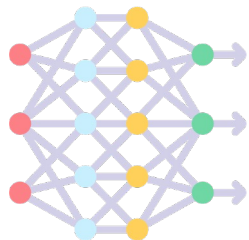
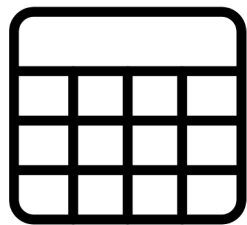
Ensemble Method	Weight	Id = 1	Id = 2	Id = 3
Hard_Voting	0	3	3	3
Weighted_Voting_1	$\left(\frac{1}{2}\right)^i / \# \text{ model}$	3.416	3.291	3.166
Weighted_Voting_2	$\frac{1}{\log_2(i + 1)}$	5.631	5.131	4.631



결과

Type	Weight	Public(30%)	Private
BM3	-	0.0359	0.0370
Hard_voting	0	0.0386	0.0399
weighted_voting_1	$\left(\frac{1}{2}\right)^i / (\# \text{ Model})$	0.0414	0.0428
weighted_voting_2	$\frac{1}{\log_2(i + 1)}$	0.0428	0.0442

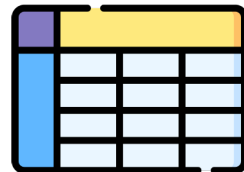
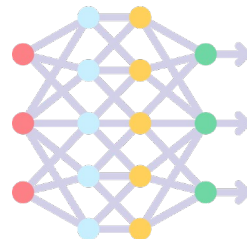
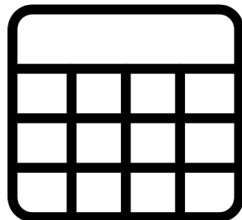
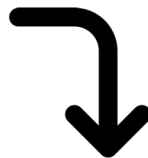
시도1 : submission 후처리



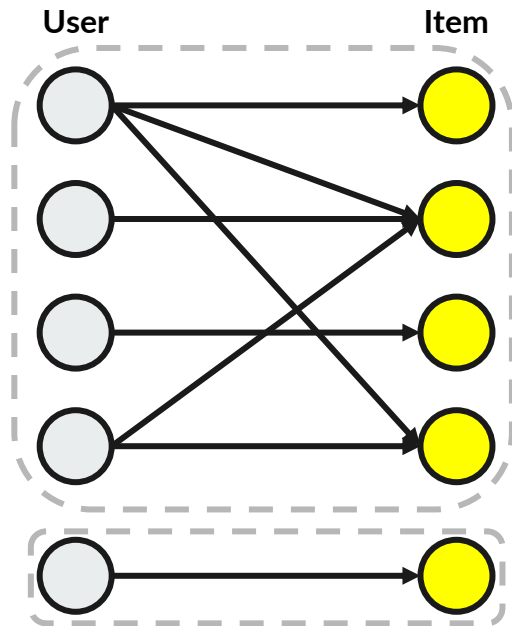
[인기상품 우선 추천]

인기상품 Top 50

- user-item 정보가 4이하인 경우
- ndcg@50 : 약 0.02 감소



시도2: 연결성분 데이터 분할



Case	Rating_All	Rating = 5
Counts	255,392	240,154
Connected	255,392	239,670

[연결그래프 분할 시도]

- Using DFS to find Connected Graph
- All node was connected
- Only Rating=5 nodes were almost connected



Q & A