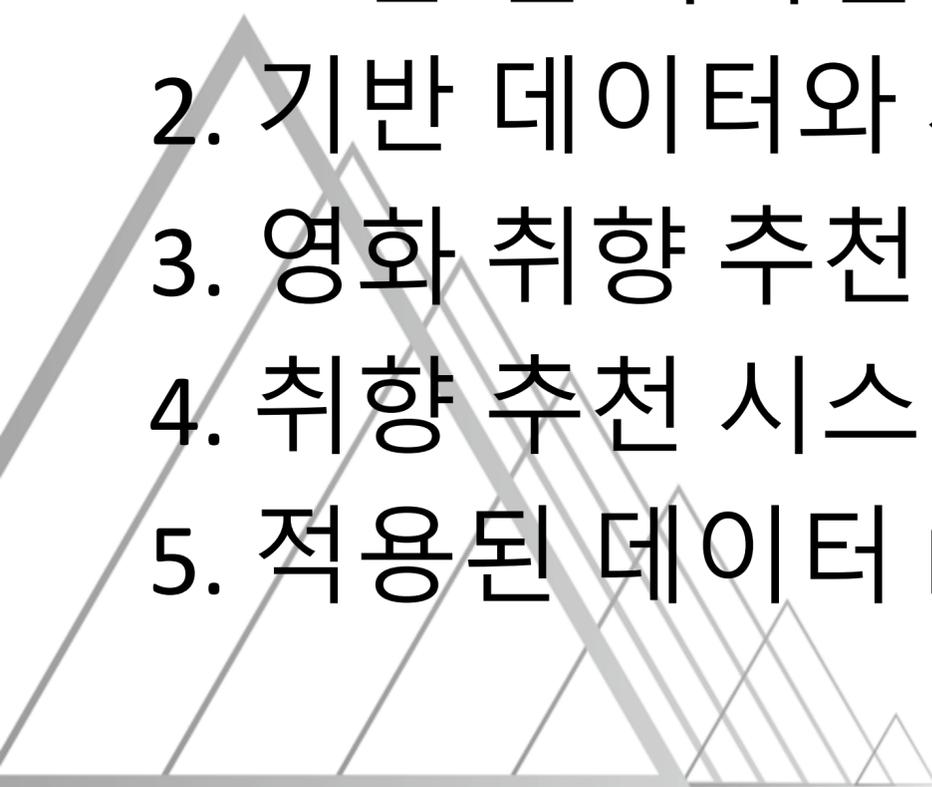


내가 선택하는 내 영화 취향: 뭘 고를지 몰라 다 준비해 봤어

권은 NAVER
김동국 NAVER

CONTENTS

1. 조금 덜 똑똑한 추천
2. 기반 데이터와 사용 알고리즘
3. 영화 취향 추천 모델
4. 취향 추천 시스템 구조
5. 적용된 데이터 Flow



1. 조금 덜 딱딱한 추천

1.1 정교한 추천, 부담스럽다

개인화 추천의 아쉬움

- 너무 정교한 추천에서 오는 거부감
- 대체로 늘 같은 종류의 추천 콘텐츠
- 변경할 수 없는 추천 주제

개인화 추천에서 오는 불만을 어떻게 해결해 줄 수 있을까?

1.2 조금 덜 똑똑한 추천

더 유연한 추천

- 개인화 추천 보다는 포괄적인 추천 제공
- 모델이 파악한 사용자 취향을 확인할 수 있게 제공
- 파악한 취향을 사용자가 수정할 수 있게 하는 기능 제공

2. 기반 데이터와 알고리즘

2.1 데이터 소개

영화 데이터

- 영화의 기본 정보 사용
 - 영화 장르
 - 제작 날짜
 - 러닝타임
 - 개봉일
 - 사용자 피드백 참여 인원

미나리
영화 | Minari | 2020

[전체](#) |
 [기본정보](#) |
 [감독/출연](#) |
 [평점](#) |
 [무비클립](#) |
 [포토](#) |
 [리뷰](#) |
 [명대사](#) |
 [보러가기](#) |
 [추천](#) >



기본정보 →

개요	드라마 미국 115분
개봉	2021.03.03.
평점	★ 8.31
관객수	113만명
수상	수상내역보기

"미나리는 어디서든 잘 자라" 낯선 미국, 아칸소로 떠난 한국 가족. 가족들에게 뭔가 해내는 걸 보여주고 싶은 아빠 '제이콥'(스티븐 연)은 자신만의 농장을 가꾸기 시작하고 엄마 '모니카'(한예리)도 다시 일자리를 찾는다. 아직 어린 아이들을 위해 '모니카'의 엄마 '순자'(윤여정)가 함께 살기로 하고 가방 가득 고춧가루, 멸치, 한약 그리고 미나리씨를 담... [더보기](#)

[보러가기](#)

♥ 5,440

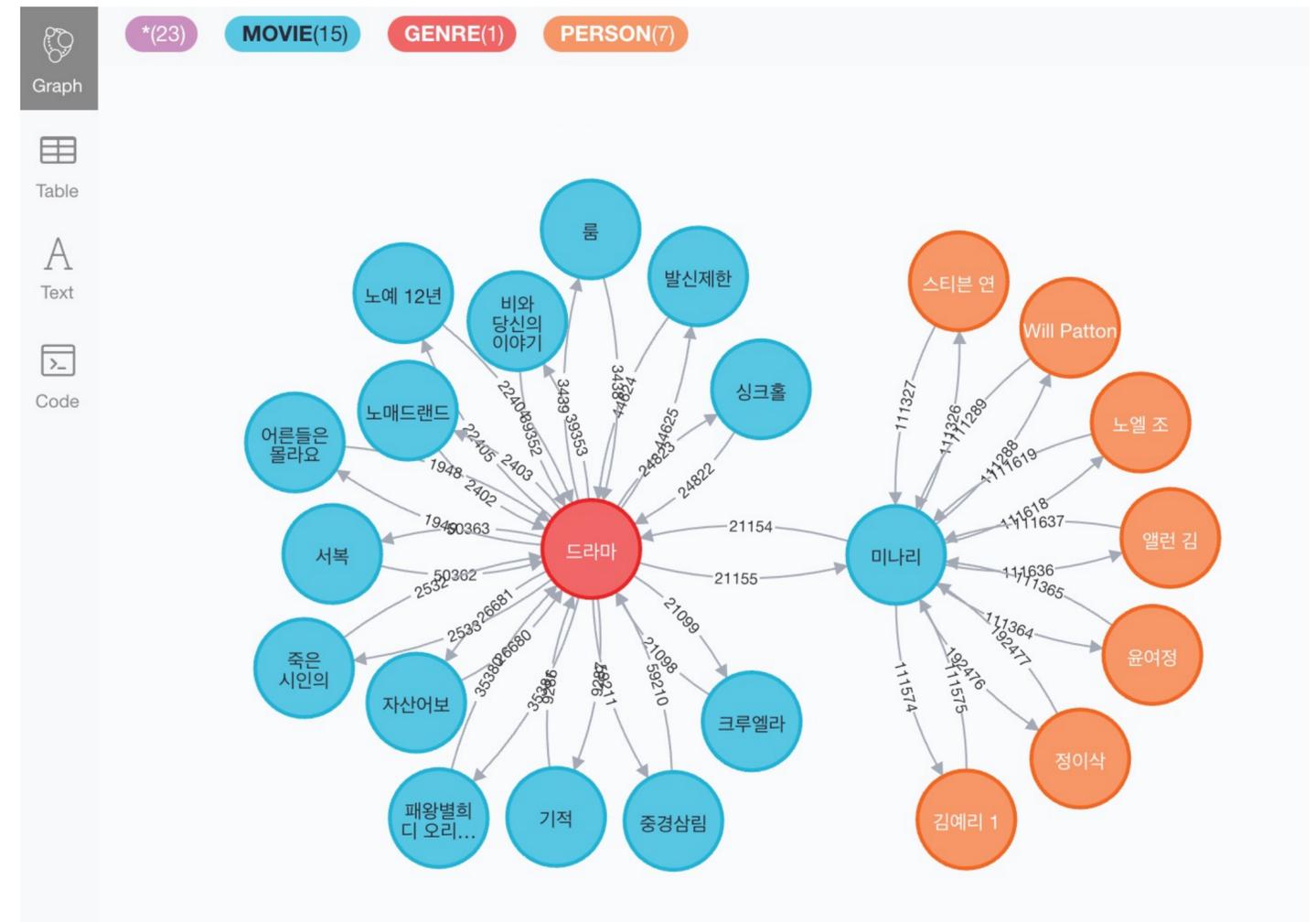
[정보오류 수정요청](#)

네이버 영화 ⓘ [다른 사이트 더보기](#)

2.2 Knowledge Graph 및 사용 알고리즘 소개

Knowledge Graph

- 정보를 기반으로 만들어진 그래프



2.2 Knowledge Graph 및 사용 알고리즘 소개

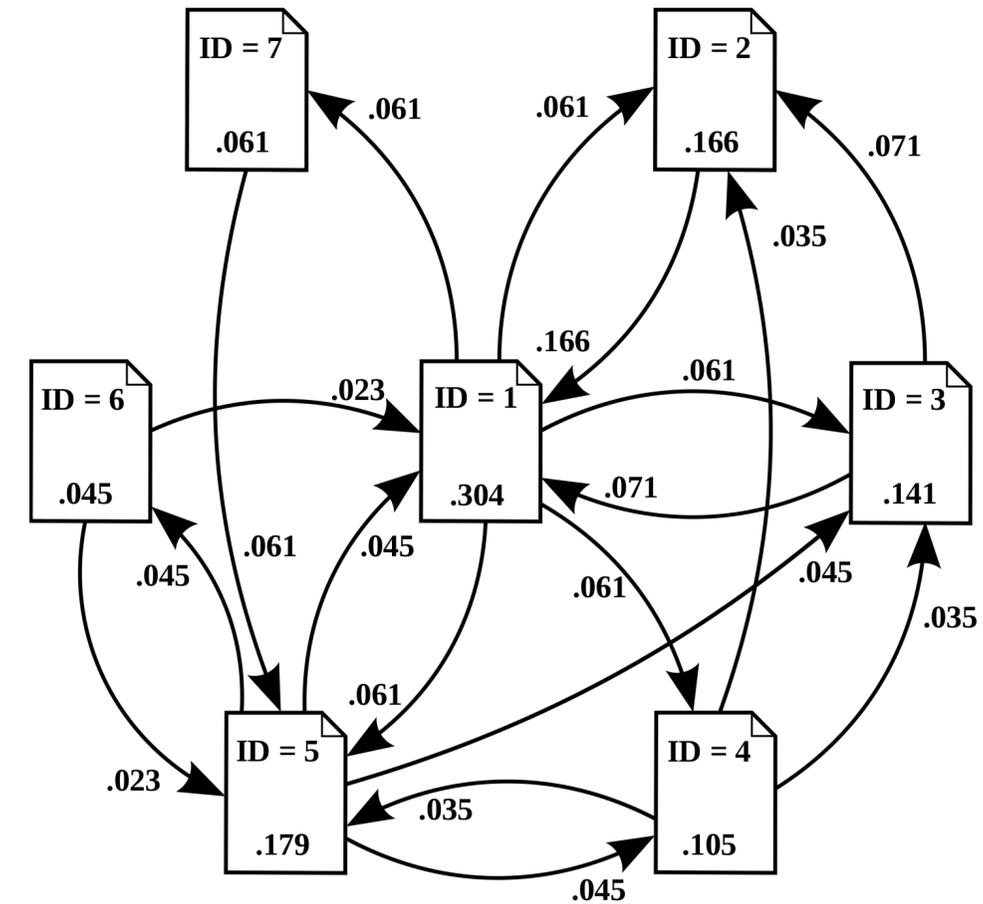
Graph 알고리즘

- Social Network Analysis Algorithm
 - Degree Centrality
 - Betweenness Centrality
 - Closeness Centrality
 - Eigen Centrality
 - **Page Rank Centrality**

2.2 Knowledge Graph 및 사용 알고리즘 소개

Page Rank

- 중요도를 연결된 Node로 나누어 가지는 방식
- 구글의 웹 사이트 페이지 중요도 측정에 쓰였던 알고리즘
- Localized 된 적용이 가능



출처 : Google wikipedia

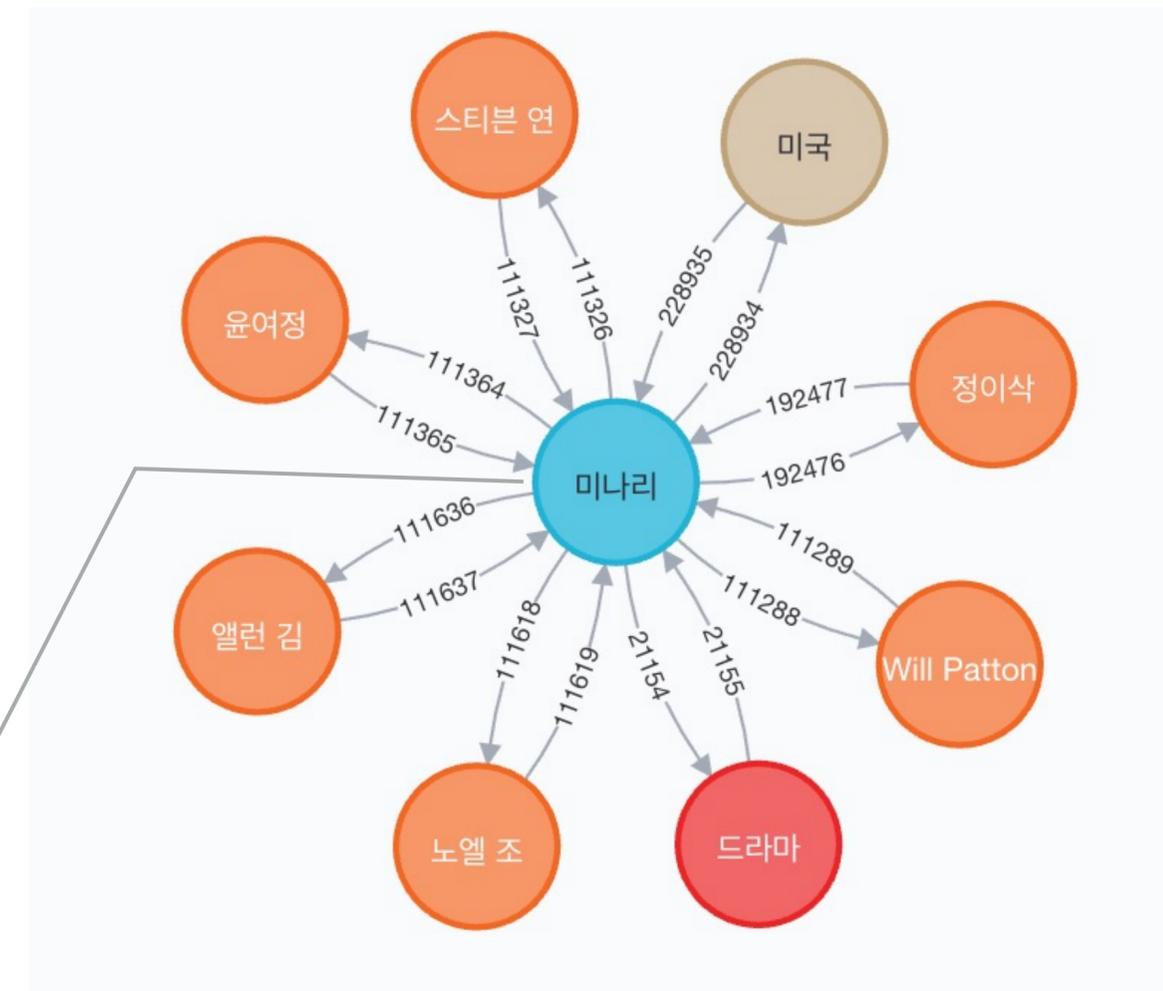
3. 영화 취향 추천 모델

3.1 Knowledge Base 구조화 및 분석

영화 데이터 Knowledge Graph

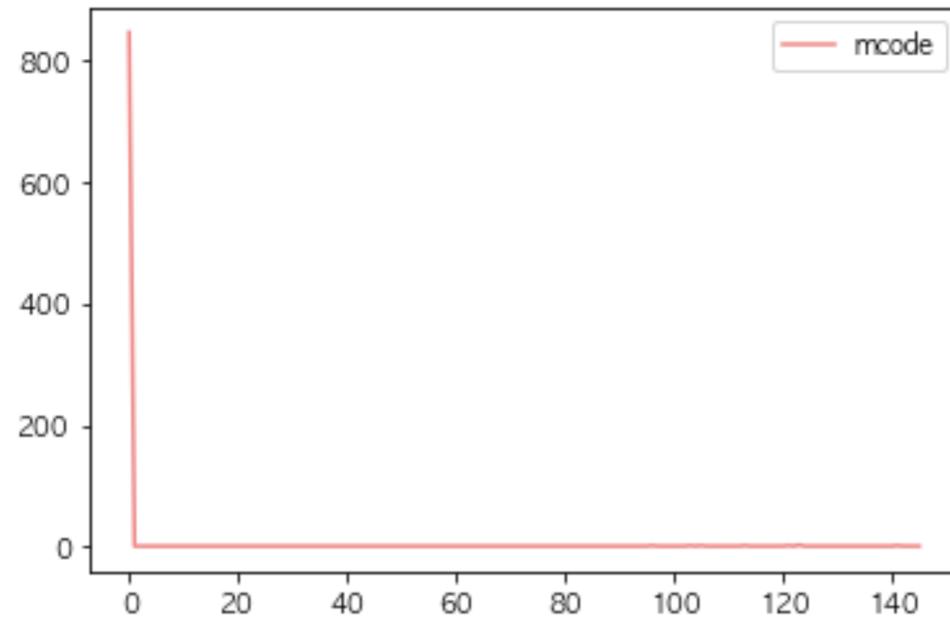
- Node: MOVIE, GENRE, NATION, PERSON
- Relation: 양방향 연결

타이틀
개봉일
사용자 피드백 참여 인원



3.1 Knowledge Base 구조화 및 분석

최신 개봉일 분석



최신 개봉일 counting line plot

	mcode
recent_release_date	
	847
20170921	3
20150820	2
20150430	2
20200813	2

최신 개봉일 counting line plot



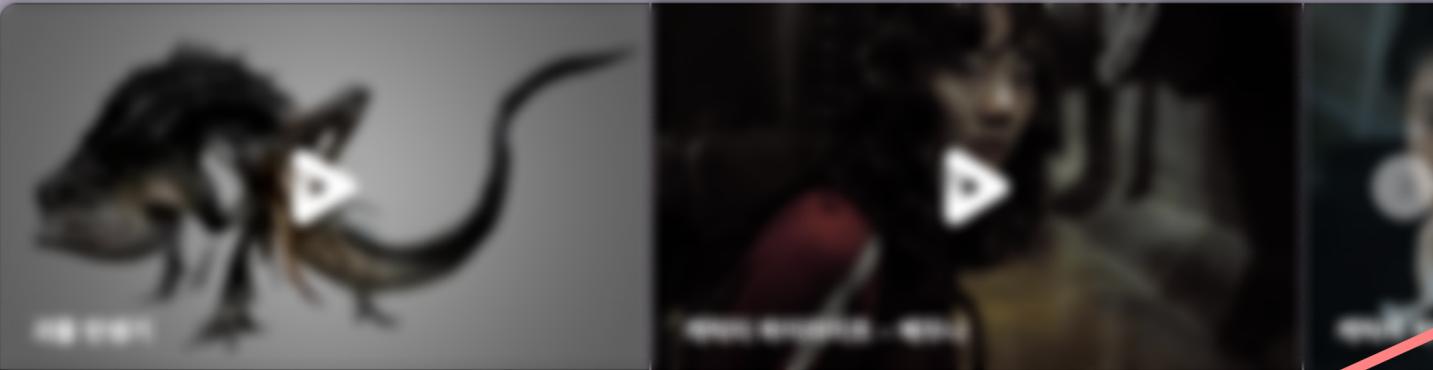
데이터 품질 및
서비스 품질을 위해
대략 85% 데이터 제거

3.1 Knowledge Base 구조화 및 분석

장르 분석

괴물
영화 | The Host | 2006

전체 기본정보 감독/출연 평점 무비클립 포토 리뷰 명대사 시리즈 보러가기 >



기본정보

개요	모험 한국 119분
개봉	2006.07.27.
평점	★ 8.62
관객수	1,091만명
수상	수상내역보기

→ Main Genre

괴물
영화 | The Host | 2006

전체 **기본정보** 감독/출연 평점 무비클립 포토 리뷰 명대사 시리즈 보러가기 >

개봉	2006.07.27.
등급	12세 관람가
장르	모험, 액션, 스릴러, 코미디, 드라마, SF, 판타지
국가	한국
러닝타임	119분
배급	(주)쇼박스

→ Sub Genre

보러가기 2,744

정보오류 수정요청

네이버 영화 다른 사이트 더보기

3.1 Knowledge Base 구조화 및 분석

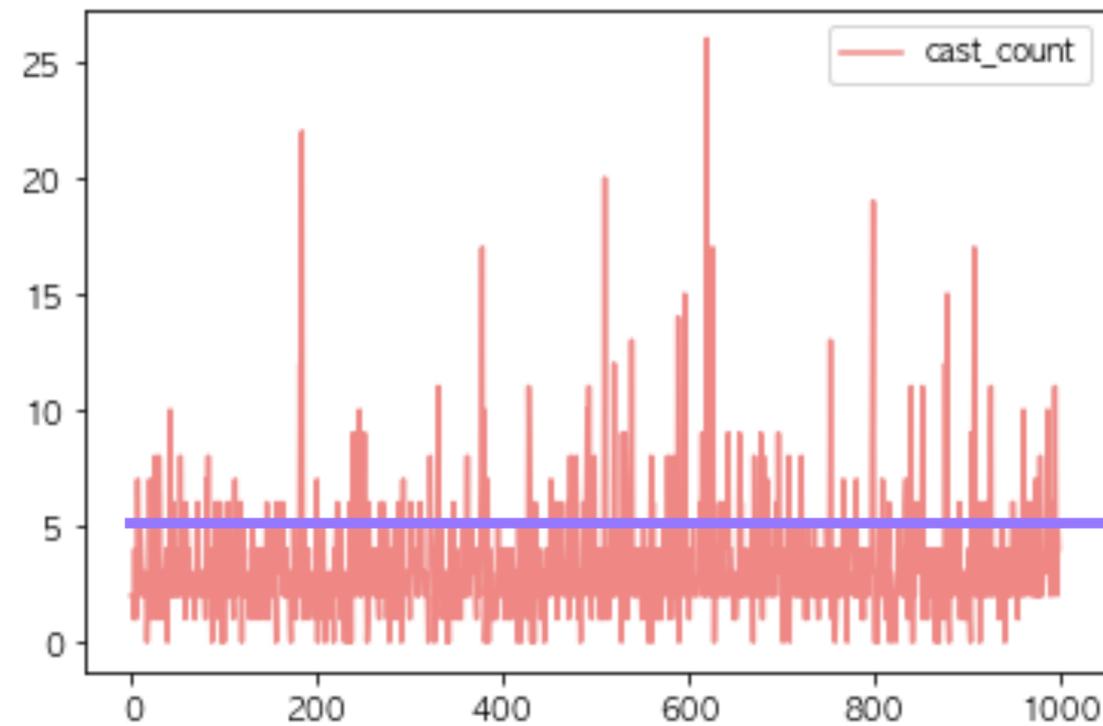
장르 분석

main_genre	드라마	액션	코미디	멜로/로맨스	애니메이션	다큐멘터리	공포	범죄	스릴러	에로	모험	미스터리	SF	공연실황	전쟁	가족	판타지	느와르	뮤지컬	무협		
count	340	122	95	89	69	49	46	37	31	21	21	16	14	14	12	11	8	2	2	1		
sub_genre	드라마	스릴러	멜로/로맨스	코미디	액션	판타지	범죄	모험	가족	SF	미스터리	공포	뮤지컬	전쟁	애니메이션	서부	무협	서스펜스	느와르	에로	다큐멘터리	공연실황
count	191	134	81	70	57	52	44	40	36	35	34	21	11	9	7	3	2	2	2	2	1	1
total_genre	드라마	액션	멜로/로맨스	코미디	스릴러	범죄	애니메이션	공포	모험	판타지	미스터리	다큐멘터리	SF	가족	에로	전쟁	공연실황	뮤지컬	느와르	서부	무협	서스펜스
count	531	179	170	165	165	81	76	67	61	60	50	50	49	47	23	21	15	13	4	3	3	2

→ 불균형한 데이터 ←

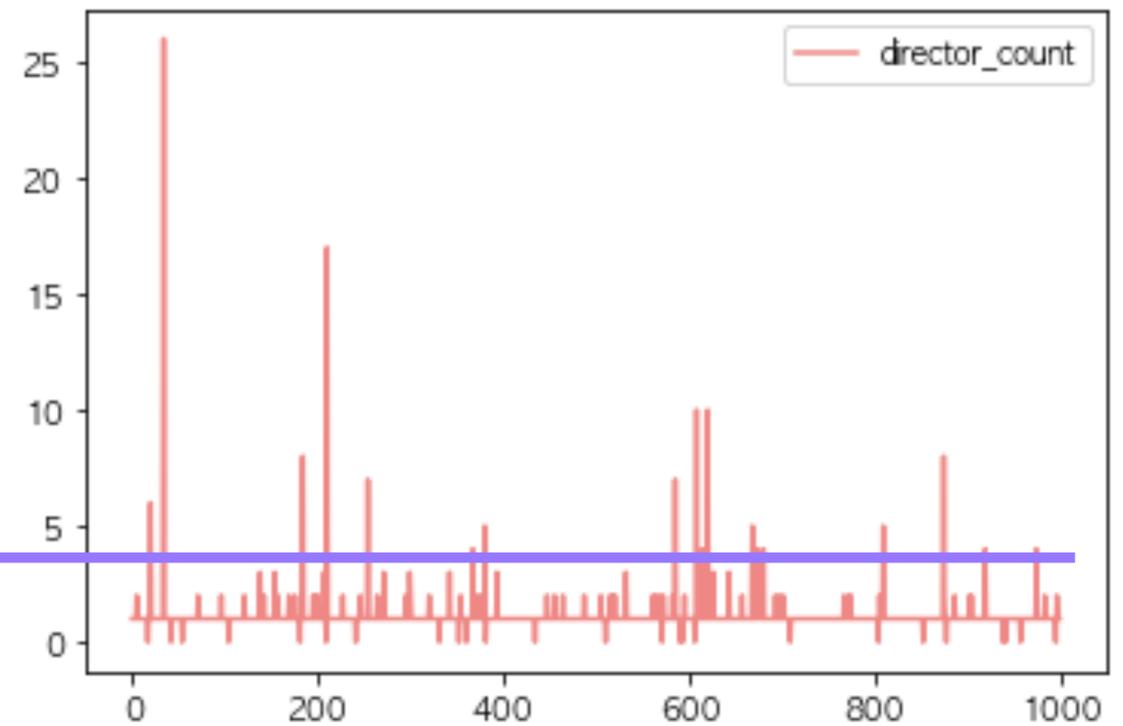
3.1 Knowledge Base 구조화 및 분석

인물 데이터 분석



한 영화에 출연한 배우 수

기준 값보다 큰
이상치 제거



영화의 참여 감독 수

3.2 영화 추천 모델 구조 제작

Needs

- 영화를 선택하면 사용자의 취향 장르가 노출 될 것
- 해당 취향 장르들의 가중치를 수정할 수 있을 것
- 취향 장르에 맞는 영화가 추천될 것

3.2 영화 추천 모델 구조 제작

FOR YOU 취향선택

라라랜드 습도 다소 높음 아임 유어 맨

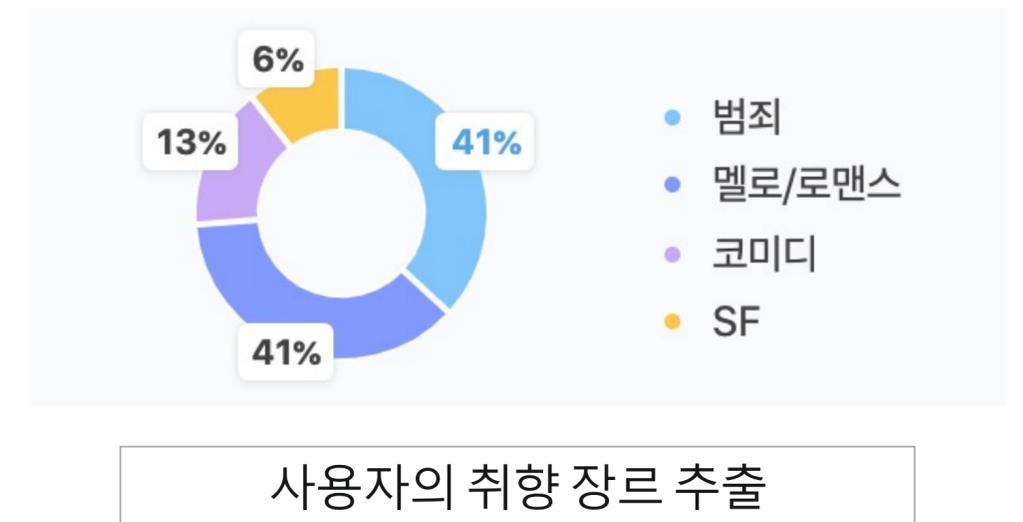
양자물리학 보이스 페르세폴리스

스피노사 톰 홀랜드 엔트로피 OLD

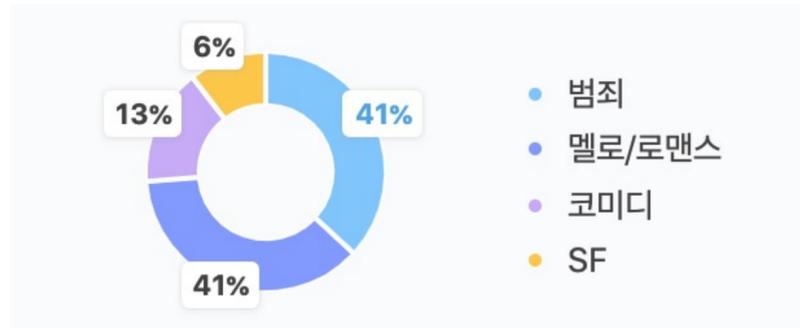
사용자가 영화 선택

취향 추출 모델

입력된 영화에 대해 Page Rank 적용

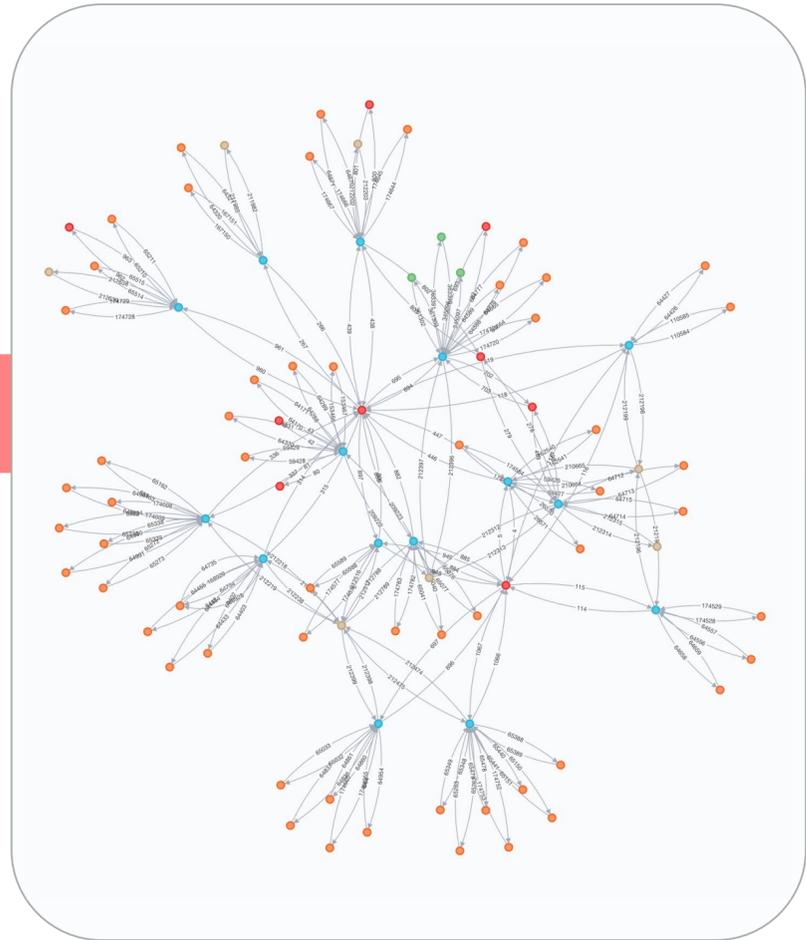


3.2 영화 추천 모델 구조 제작



사용자의 취향 장르

추천 영화 추출 모델



입력된 장르에 대해 Page Rank 적용

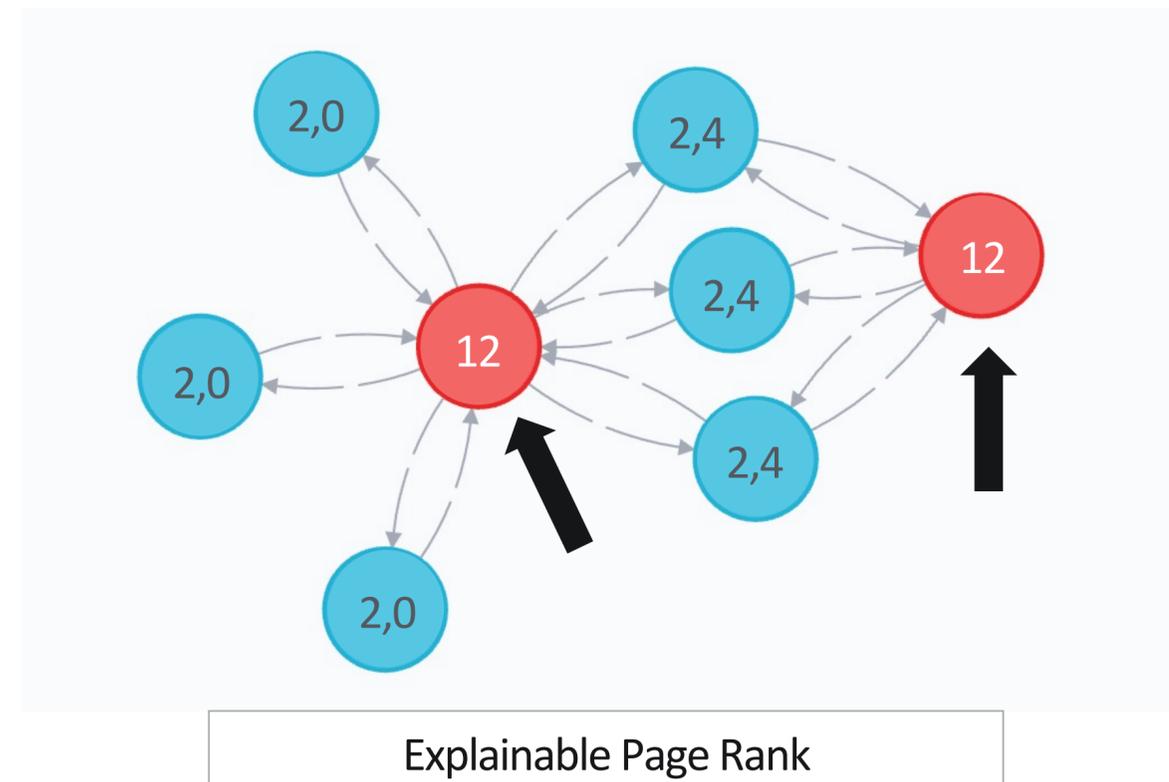
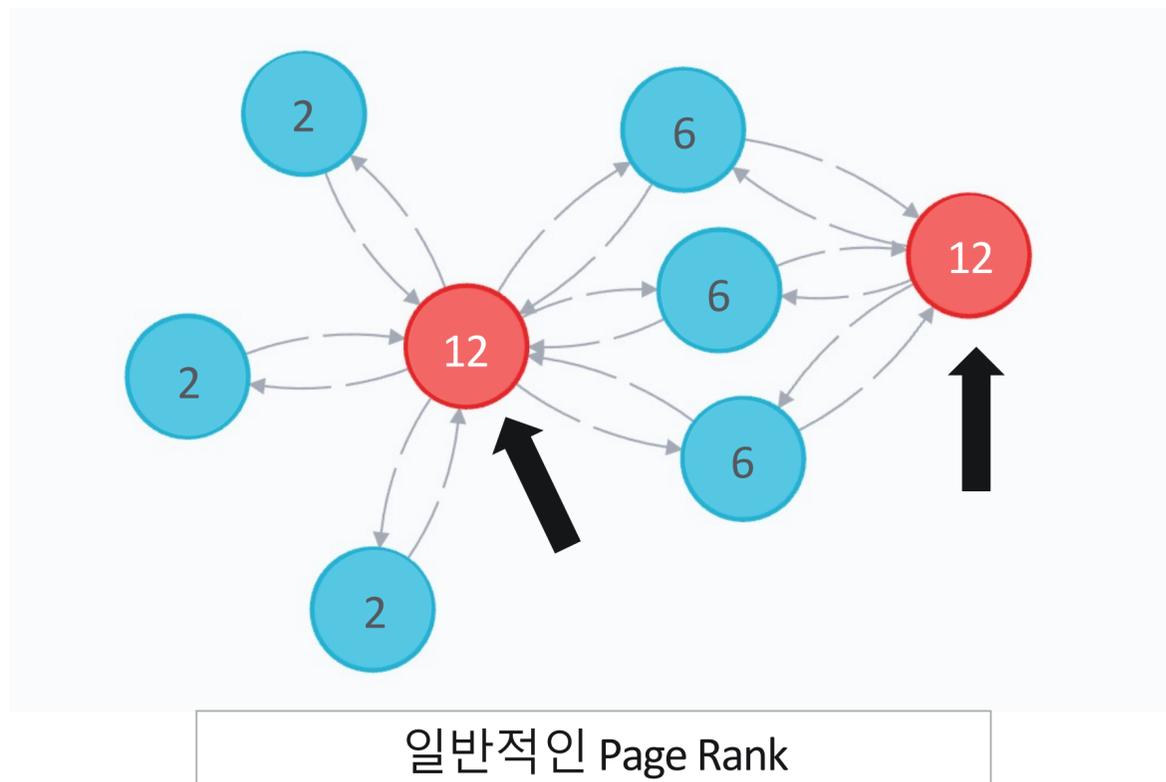
영화 제목	평점
소리도 없이	★ 7.81
암수살인	★ 8.58
블랙머니	★ 8.75
골든슬럼버	★ 6.26
범털	★ 6.26
그 시절, 우리가 좋아했던 소녀	★ 9.14

취향 장르에 알맞는 영화 추천

3.3 영화 추천 모델 제작 및 개선

Explainable Page Rank

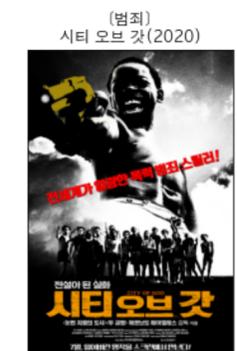
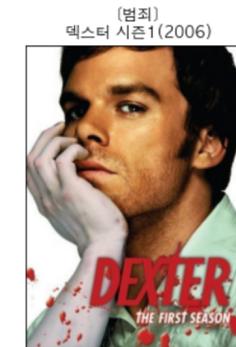
- 취향 장르 가중치 비율을 조정했을 때 추천 영화도 변경되어야 한다
- Page rank 점수를 입력 받는 취향 장르 별로 기록



3.3 영화 추천 모델 제작 및 개선

Page Rank만 적용한 결과

- 난해한 영화 추천
- 새로운 영화가 추가되어도 변화가 없는 결과
- 대중성, 최신성 반영이 필요



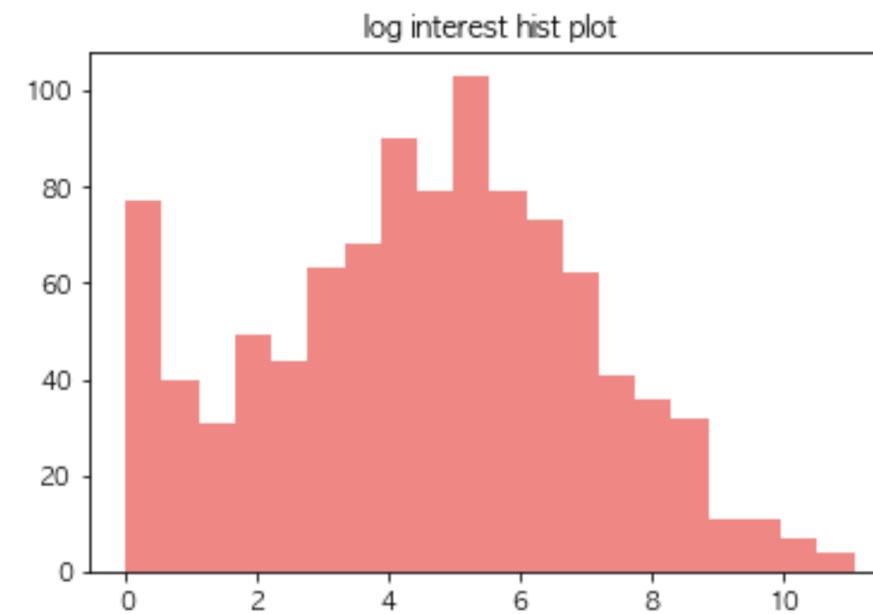
3.3 영화 추천 모델 제작 및 개선

관심도 반영

- 대중성 반영
- 감정과는 관련없이, 충분히 관심을 받은 작품인가?를 파악하는 것이 목표

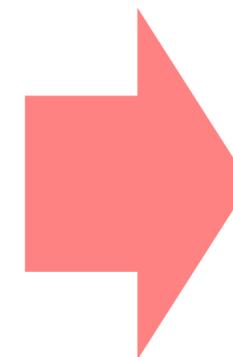
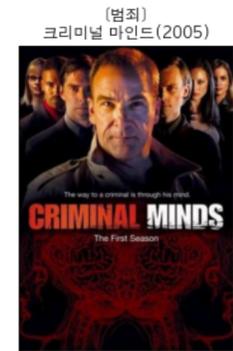
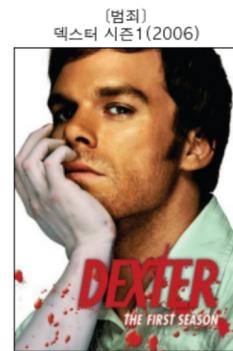


Log scale



3.3 영화 추천 모델 제작 및 개선

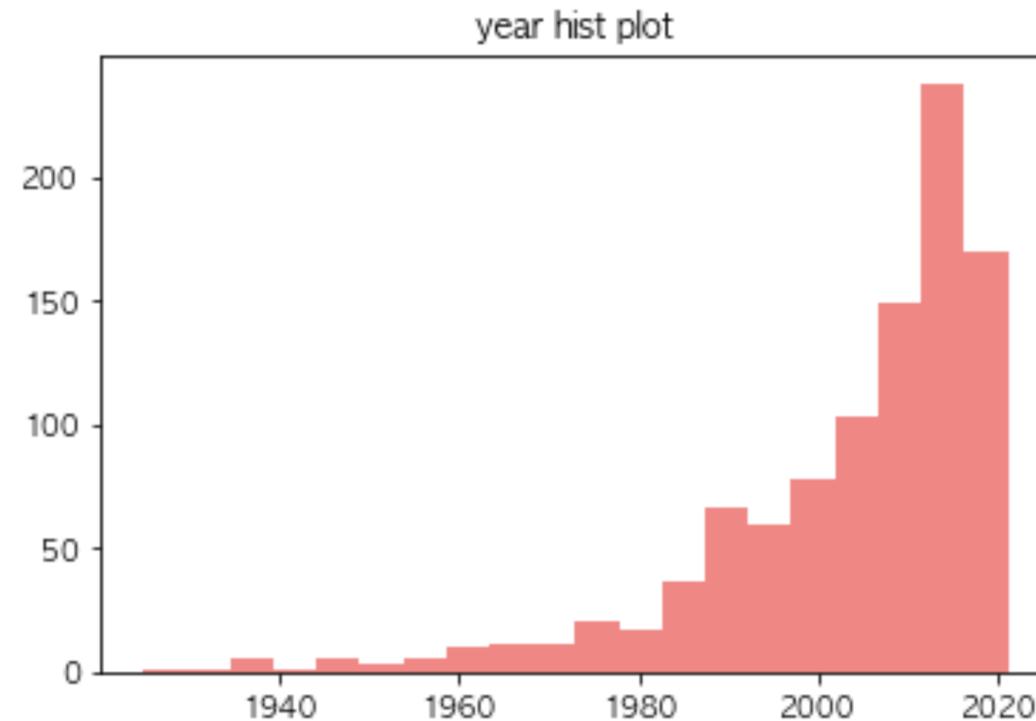
영화 추천 결과



3.3 영화 추천 모델 제작 및 개선

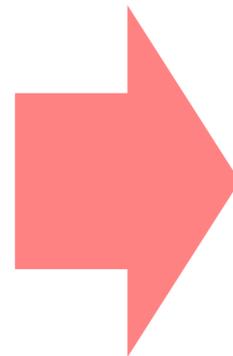
최신성 반영

- 개봉년도 데이터를 적용해 포괄적으로 최신성 적용
- 최근 개봉 영화와 관련된 needs 만족을 위해 최근 개봉일 데이터 적용



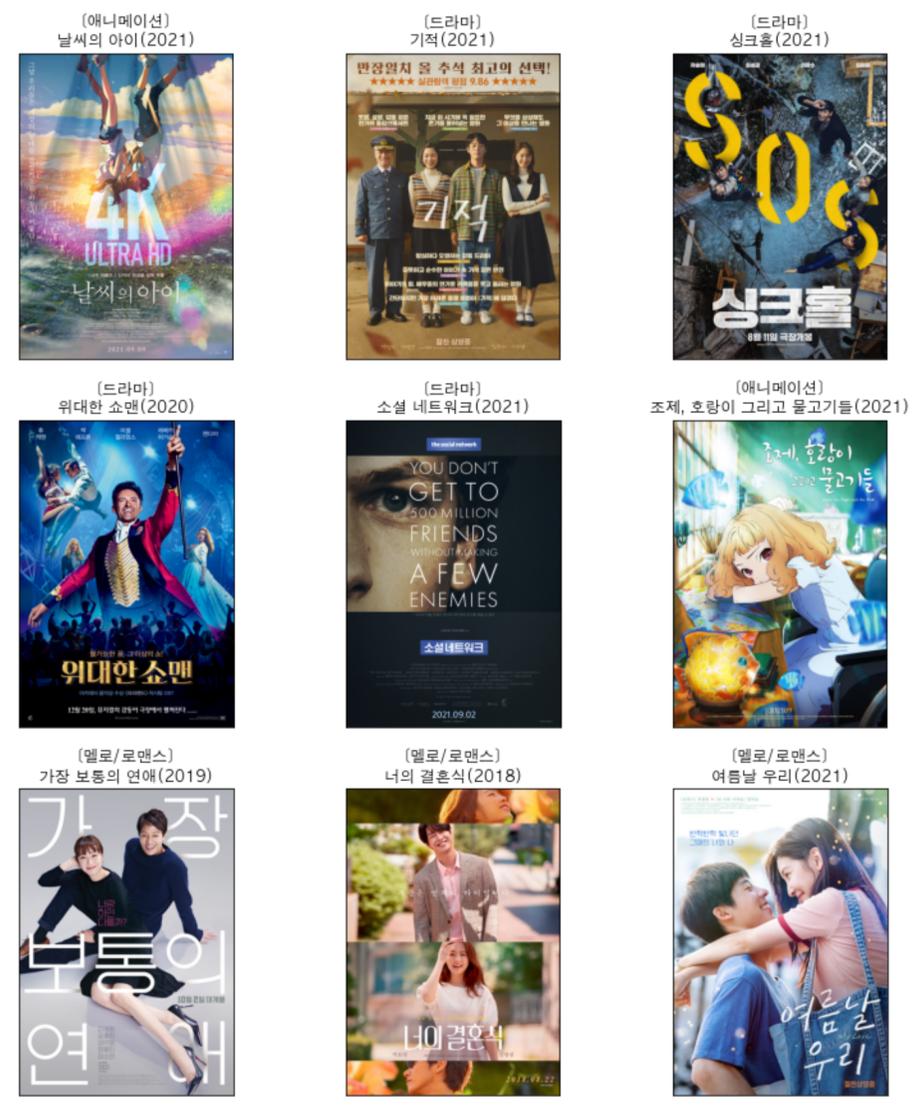
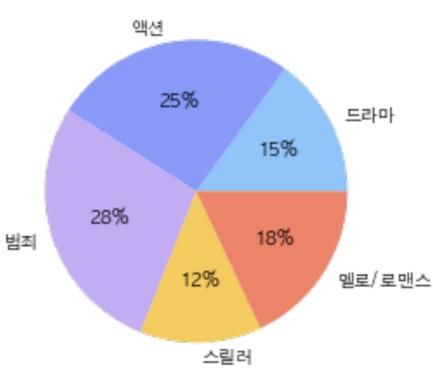
3.3 영화 추천 모델 제작 및 개선

영화 추천 결과

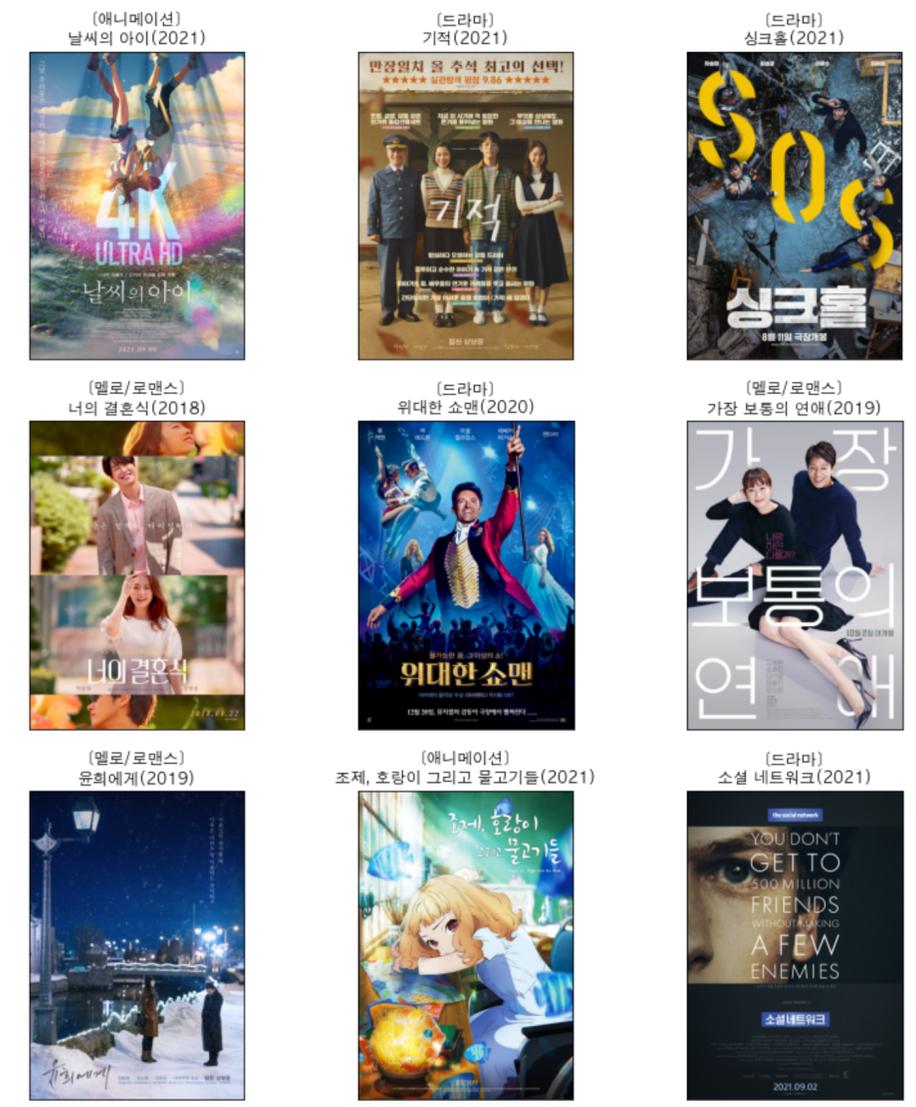


3.3 영화 추천 모델 제작 및 개선

장르 가중치 반영 시 추천 영화의 변화



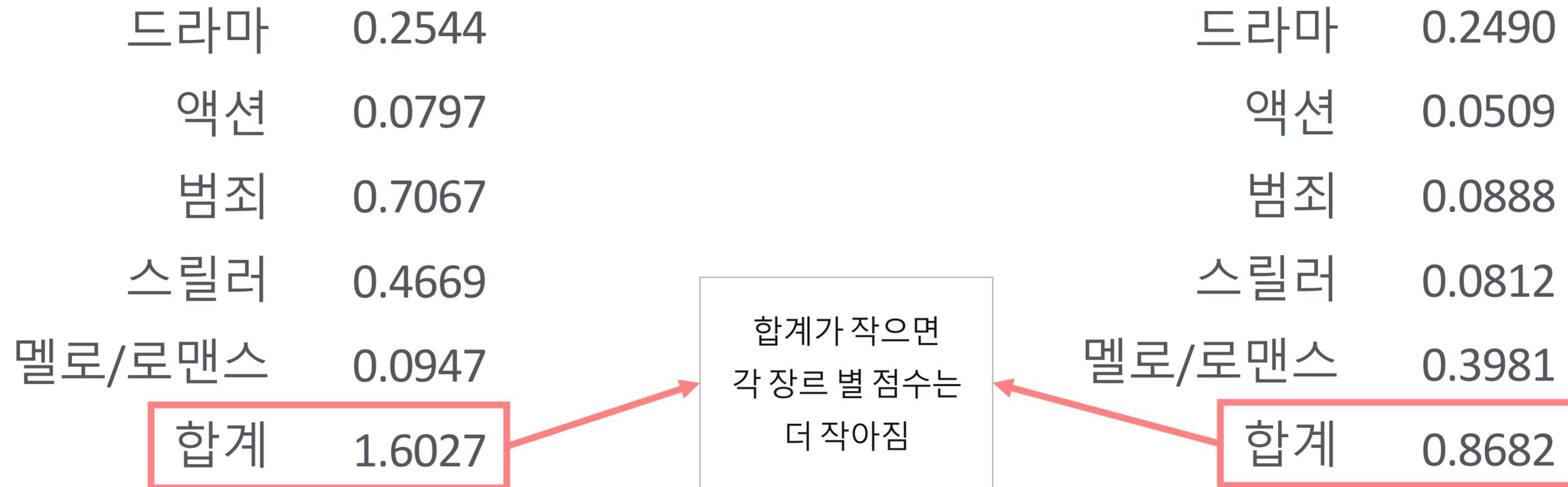
유사함



3.3 영화 추천 모델 제작 및 개선

장르 가중치 반영이 잘 작동하지 않는 이유

- 합계 점수의 영향력이 과도하게 큼



3.3 영화 추천 모델 제작 및 개선

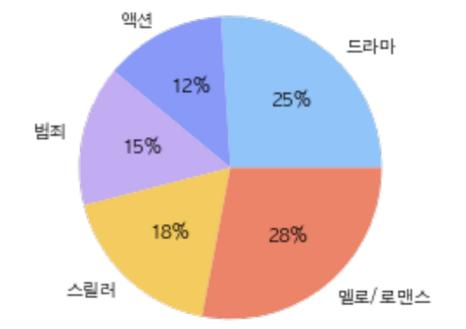
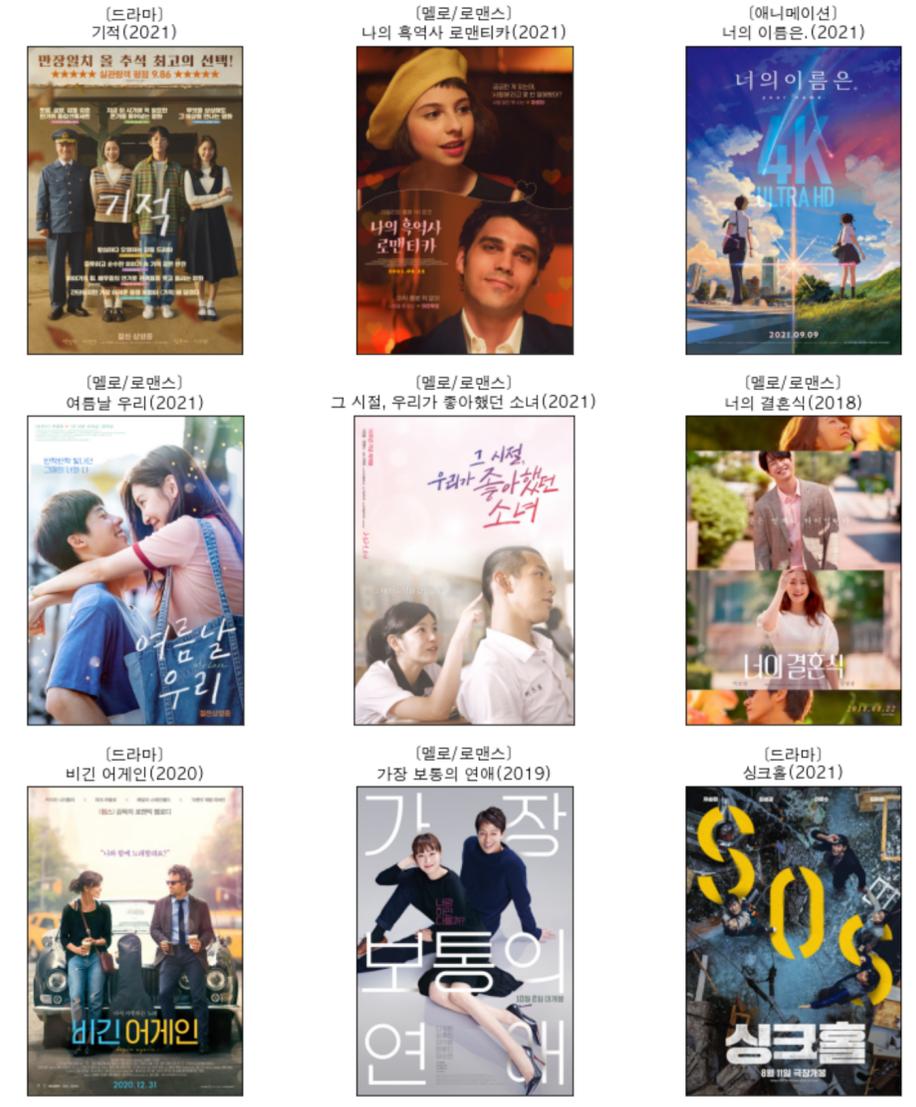
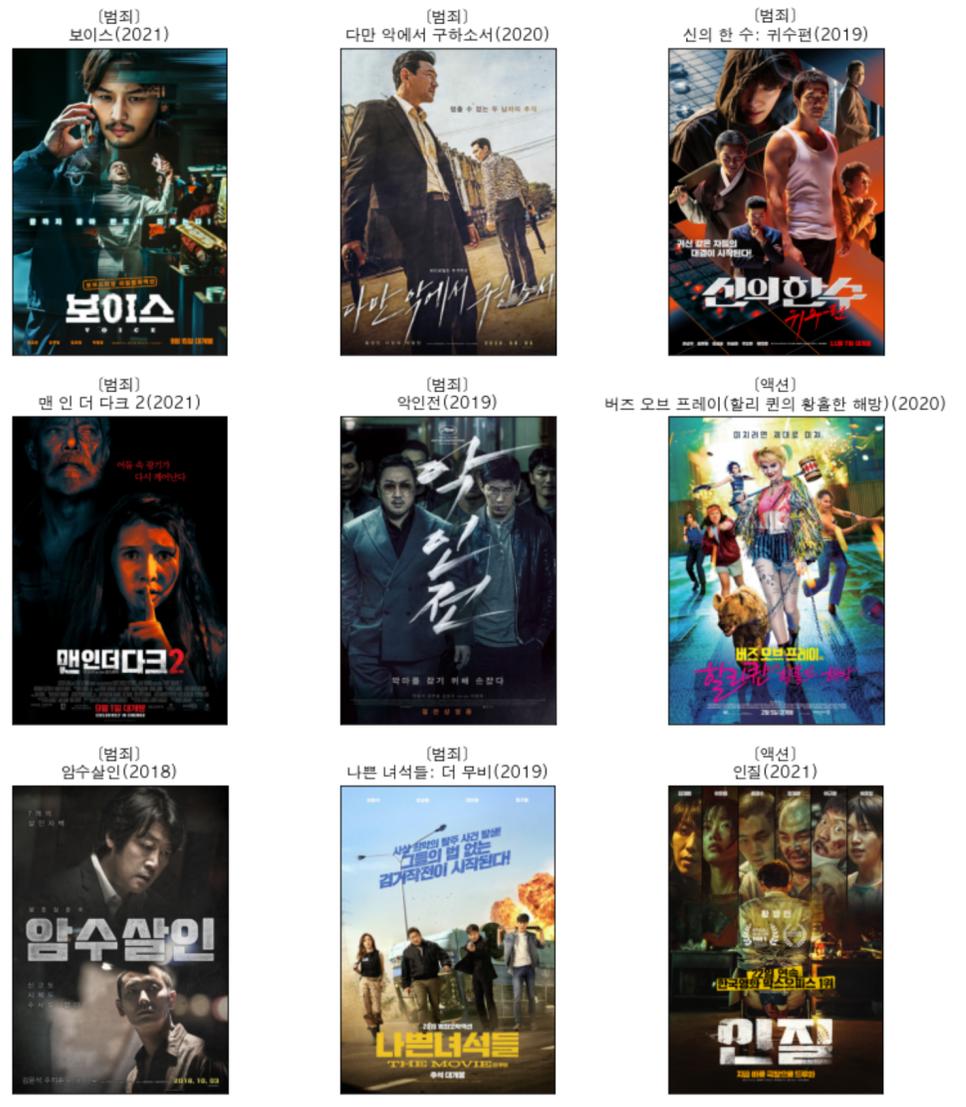
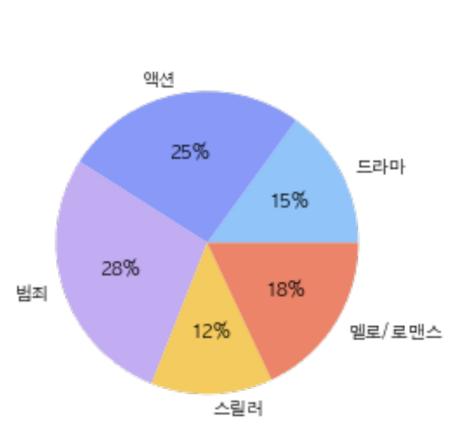
영화 별 장르 선호도 비율화

- 각 영화에서 장르의 영향력을 비율화 하여 사용자 가중치 반영

드라마	0.2490	→	드라마	28%
액션	0.0509		액션	5%
범죄	0.0888		범죄	10%
스릴러	0.0812		스릴러	9%
멜로/로맨스	0.3981		멜로/로맨스	45%
합계	0.8682		합계	0.8682

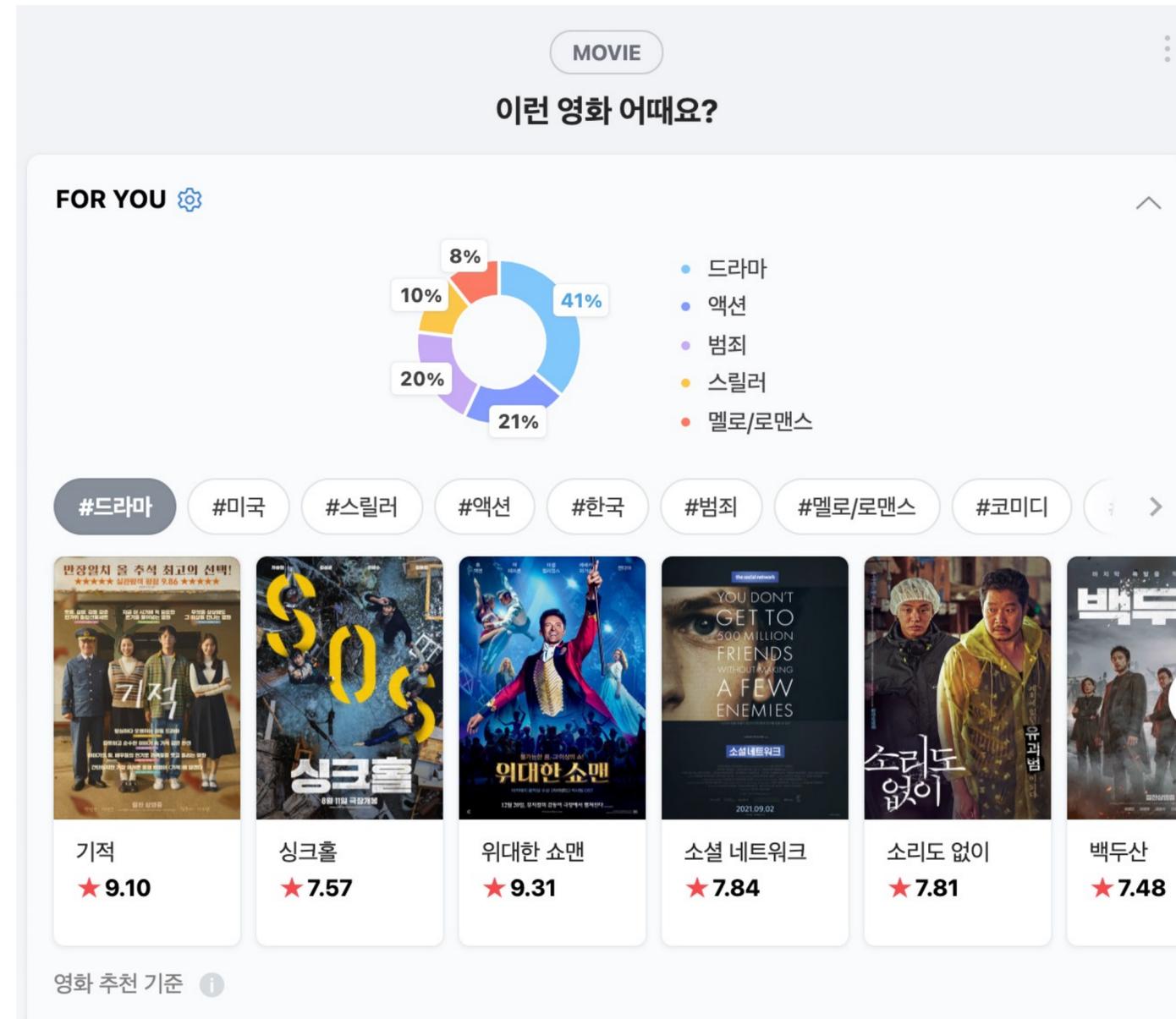
3.3 영화 추천 모델 제작 및 개선

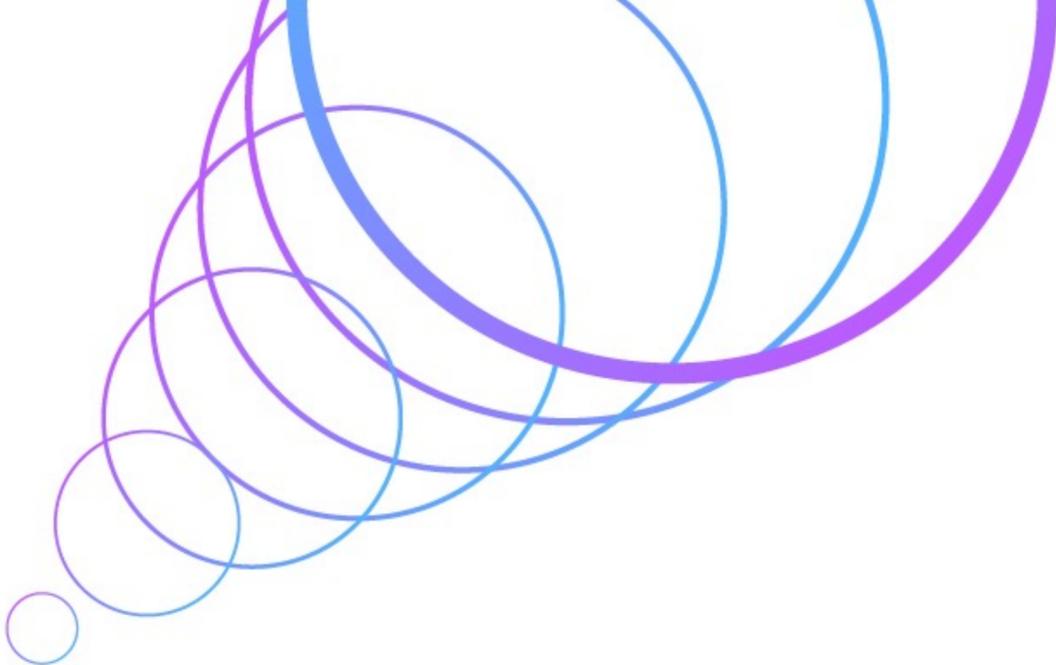
장르 가중치 반영 시 추천 영화의 변화



3.3 영화 추천 모델 제작 및 개선

결과





내가 선택하는 내 영화 취향: 뭘 고를지 몰라 다 준비해 봤어

권은 NAVER
김동국 NAVER

4. 취향 추천 시스템 구조

4.1 모델의 연산 시간

필요한 연산의 종류

- 선택한 영화에서 취향을
- 취향에서 추천할 영화를
- 사용자 별 취향을 변경했을 때 영화를

4.1 모델의 연산 시간

Page Rank

- 12개의 Object에서부터 계산을 시작하면
 - 취향 추출 3171ms
 - 키워드 추출 2622ms
 - 추천 영화 추출 3739ms
 - 총 10초 이상 소요

Started streaming 5 records after 3170 ms and completed after 3171 ms.

Started streaming 15 records after 2593 ms and completed after 2622 ms.

Started streaming 240 records after 3739 ms and completed after 3741 ms.

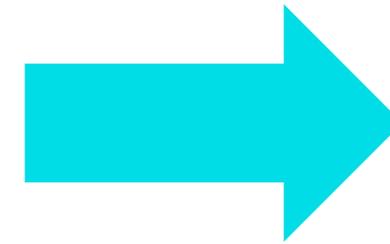
Name	Ready	Phase	TCP	Monitoring	Age	podIp	Container	CPU	MEM
ongdb-m...	Ready	Running	N/A		5 hours	10.170.48.18	ongdb-movie-dei	52.91%	30.26%

모두 실시간으로 실행하기에는 플랫폼이 버티기 어렵다

4.2 3단계 구조

항상 공통된 계산

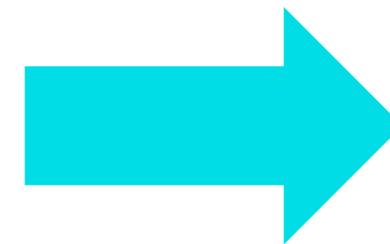
- 선택한 영화에서 취향을 뽑는 부분
- 취향에서 추천 영화를 뽑는 부분



모든 조합
정기
업데이트

별개로 분리된 계산

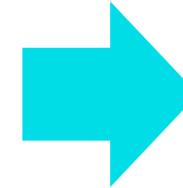
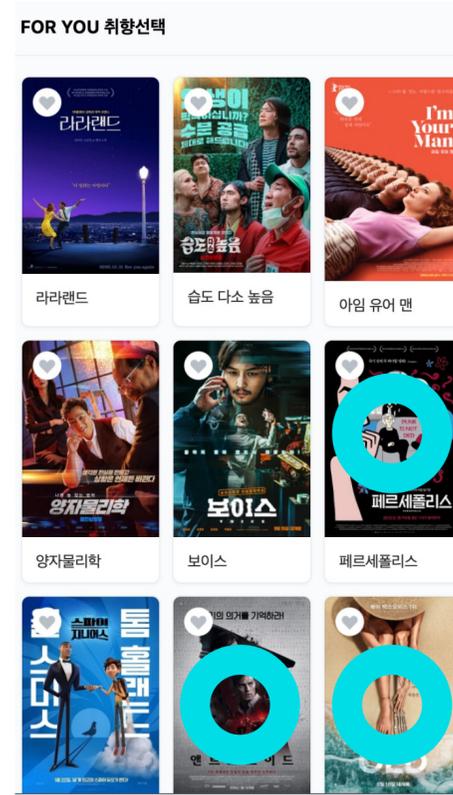
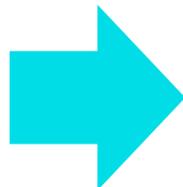
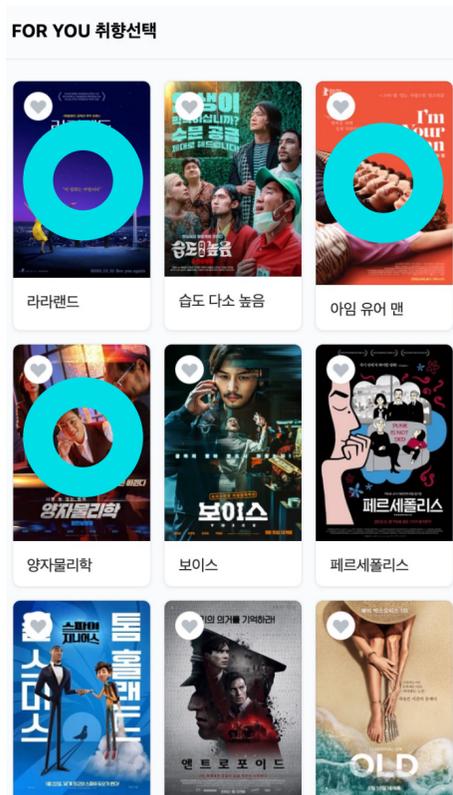
- 상세 취향 별 추천 영화 계산



점수를
저장!

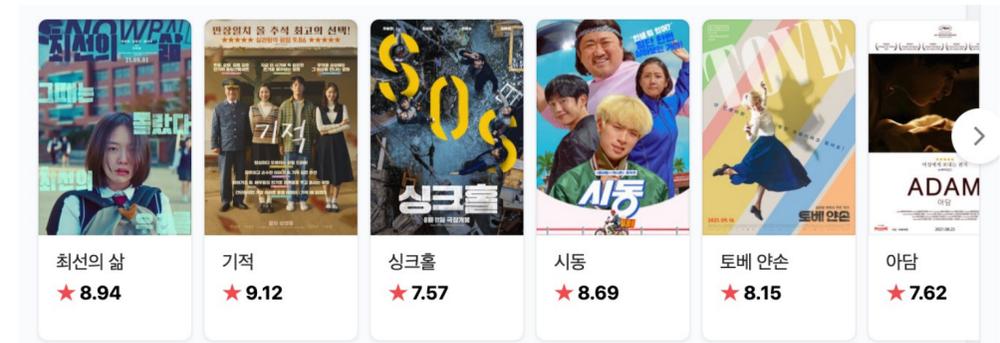
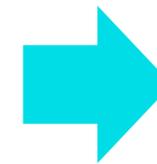
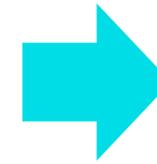
4.2 3단계 구조

1단계: 영화 -> 취향



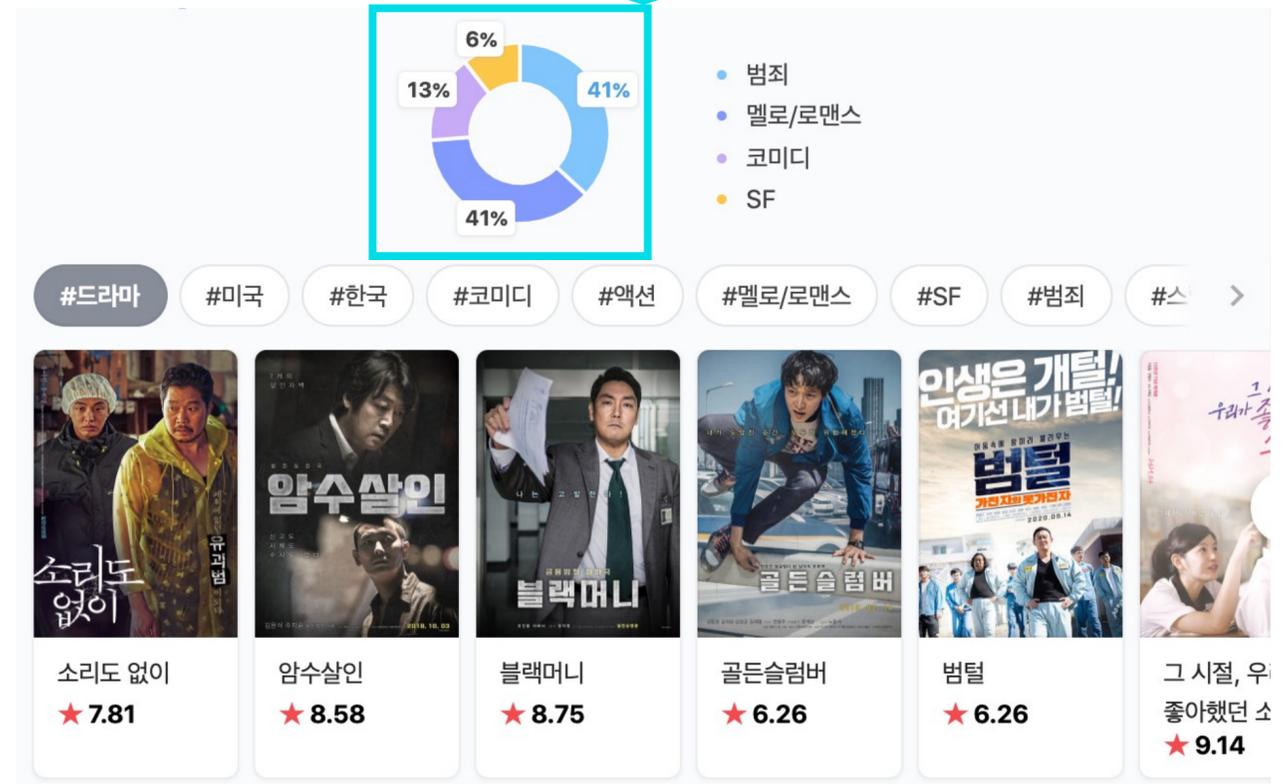
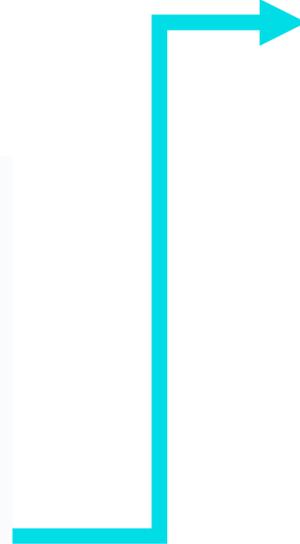
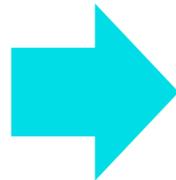
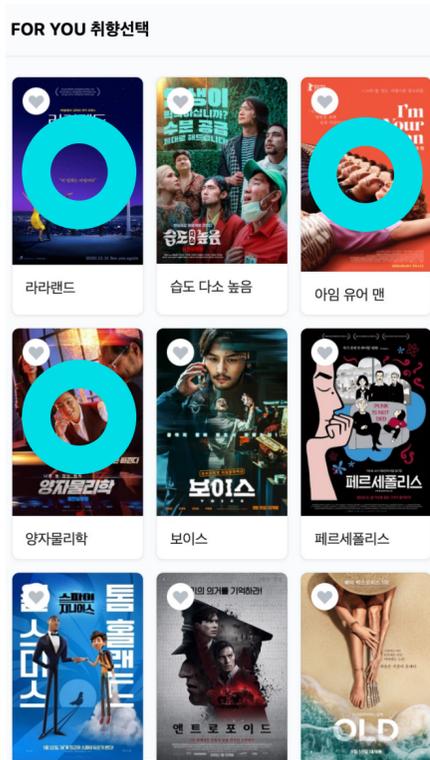
4.2 3단계 구조

2단계: 취향 장르 -> 영화



4.2 3단계 구조

3단계: 취향 추천 영화



5. 적용된 데이터 Flow

5.1 구성 요소 / 데이터 Flow

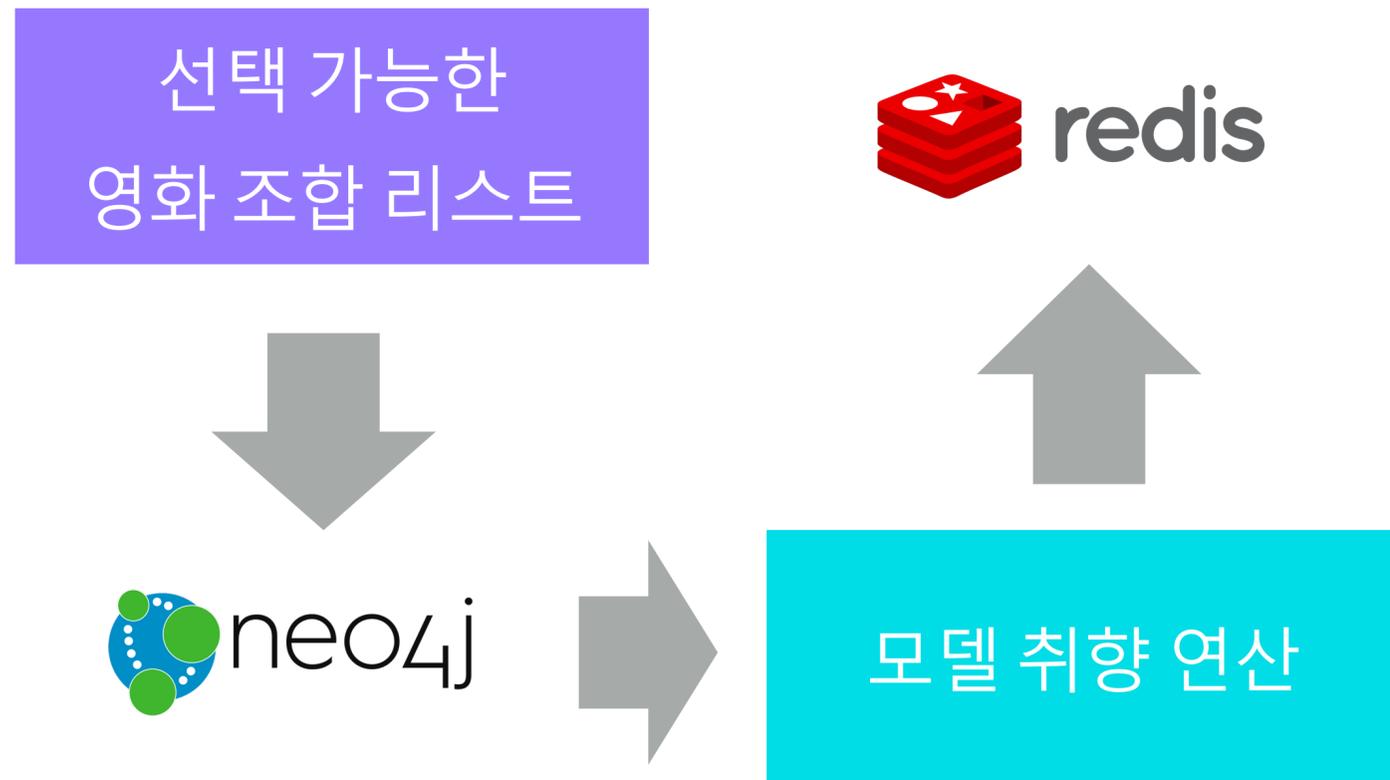
핵심요소는 GraphDB, KV Store

- 정기 데이터 흐름
- 즉각적인 데이터 흐름



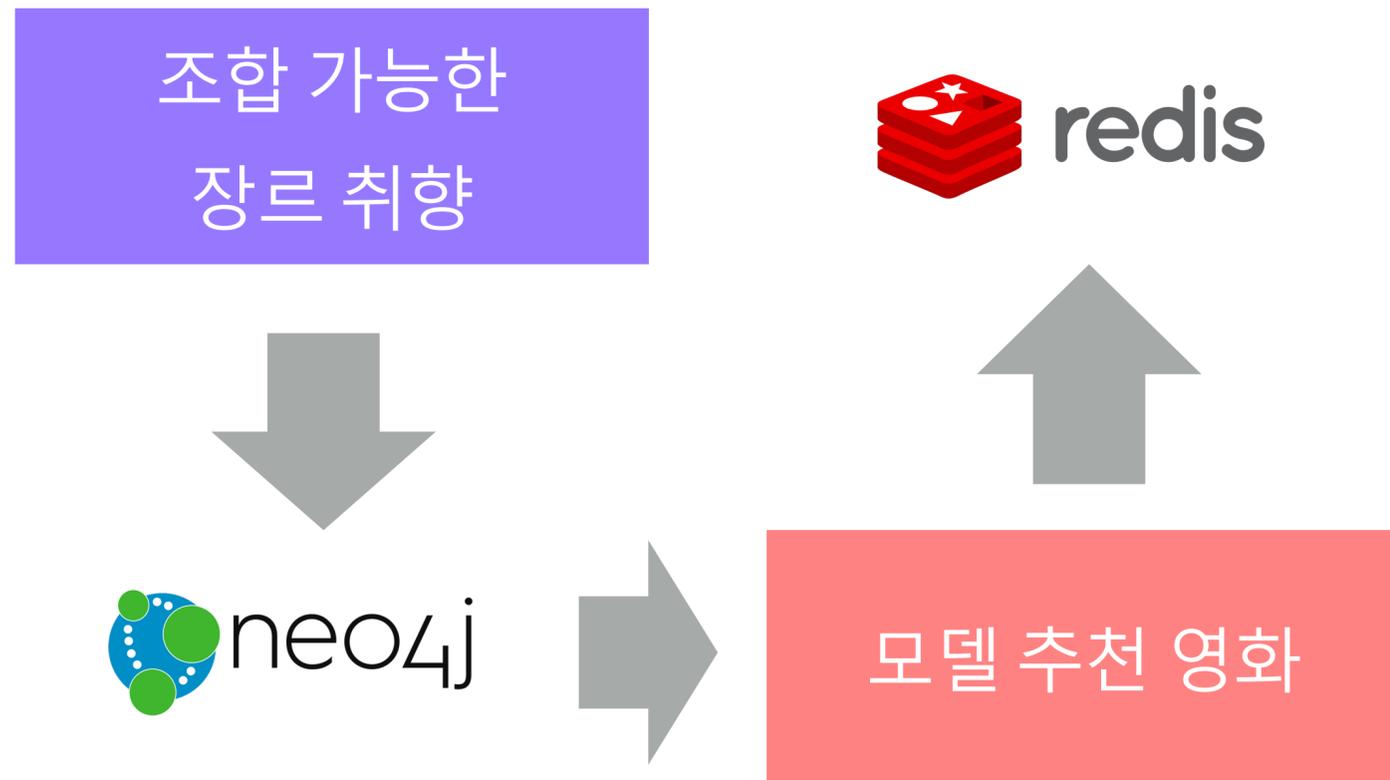
5.2 정기 업데이트

1단계 : 영화 -> 취향



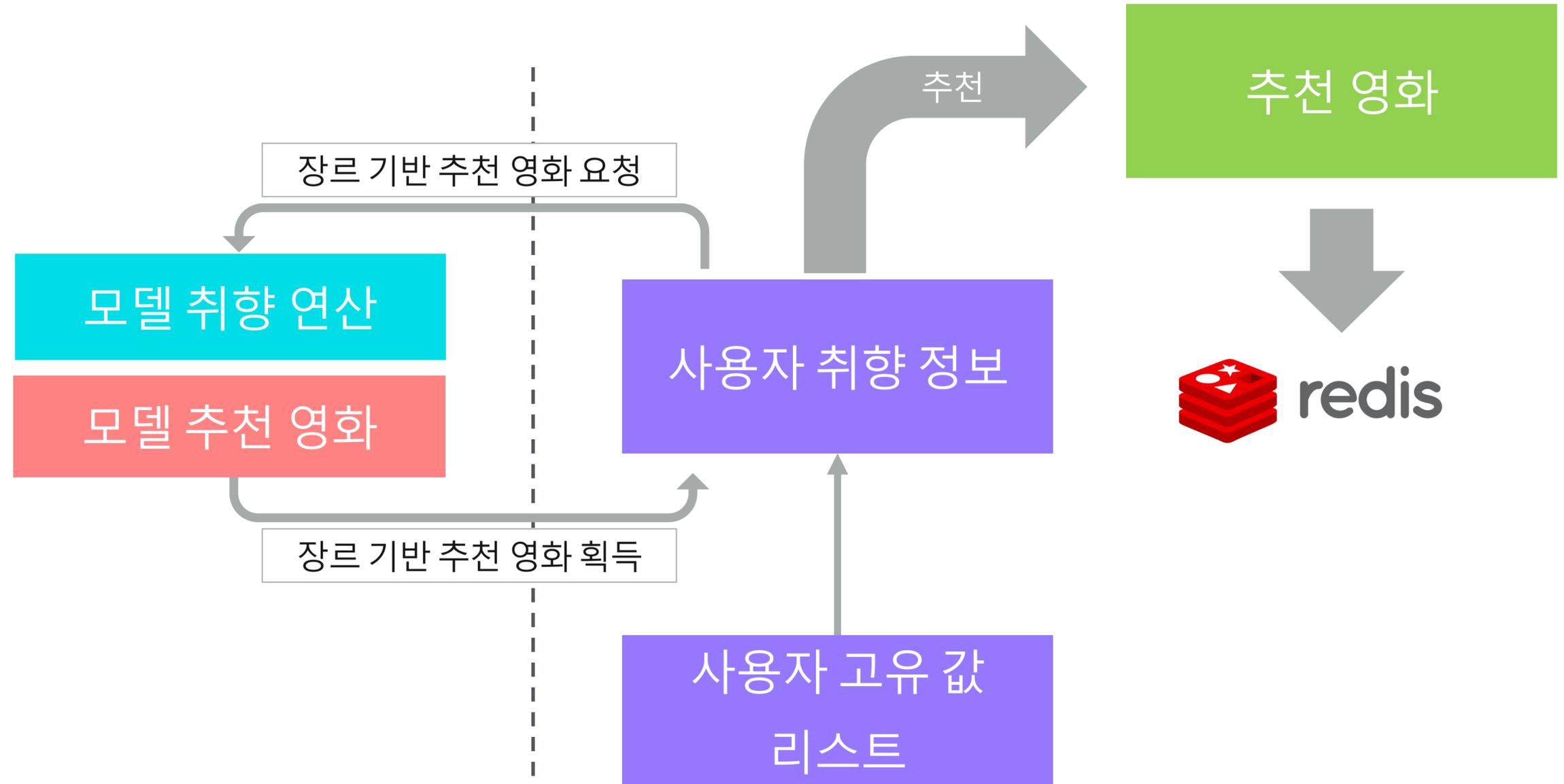
5.2 정기 업데이트

2단계 : 취향 장르 -> 영화



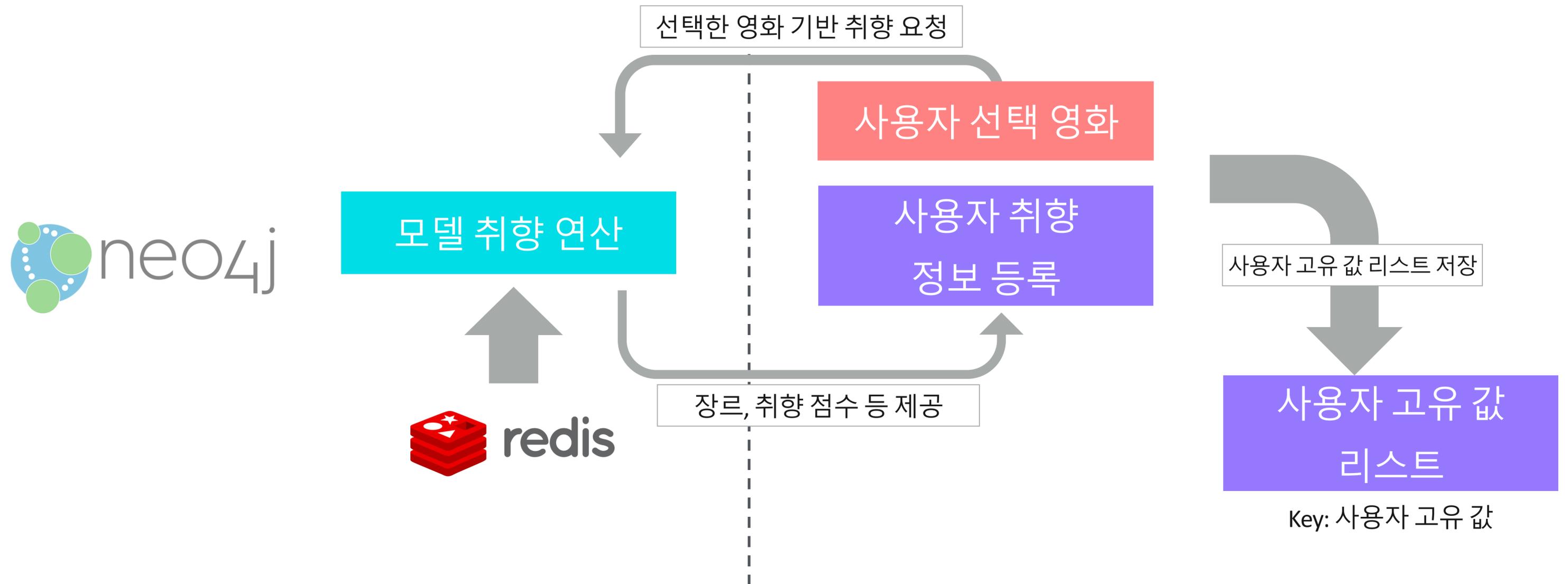
5.2 정기 업데이트

3단계 : 추천



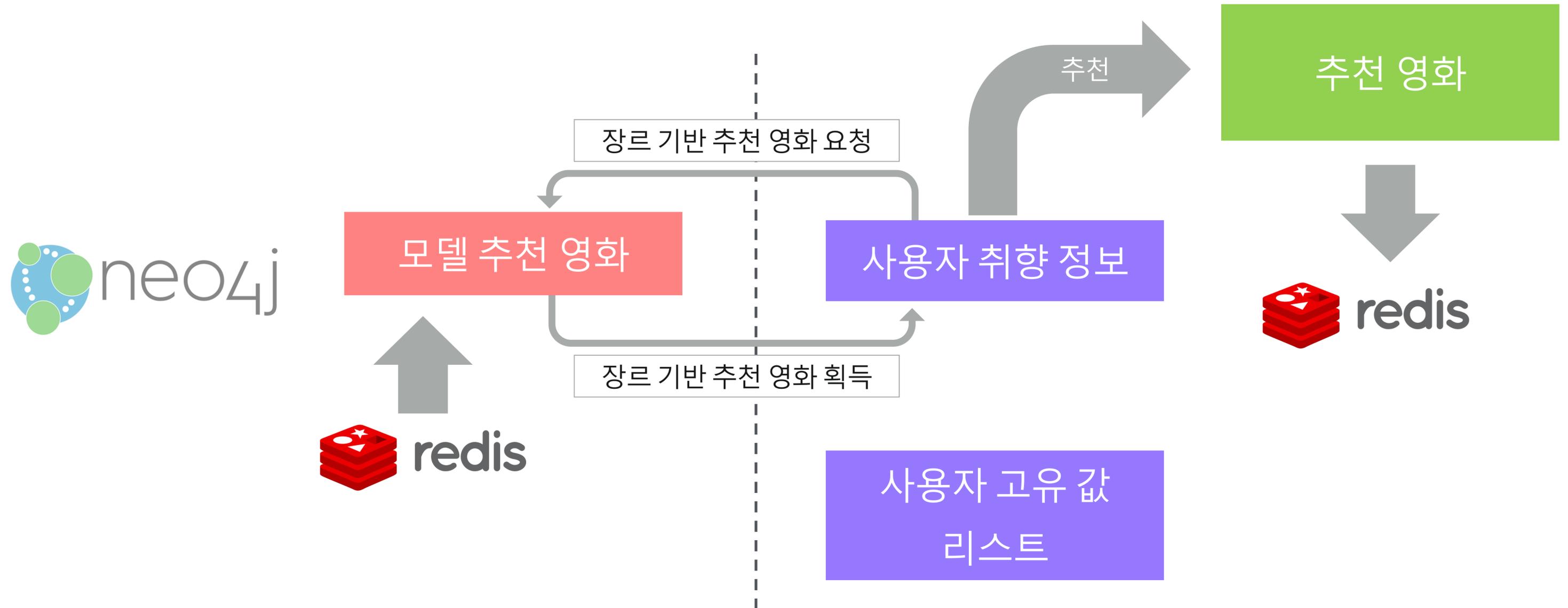
5.3 상시 업데이트

사용자 취향 획득



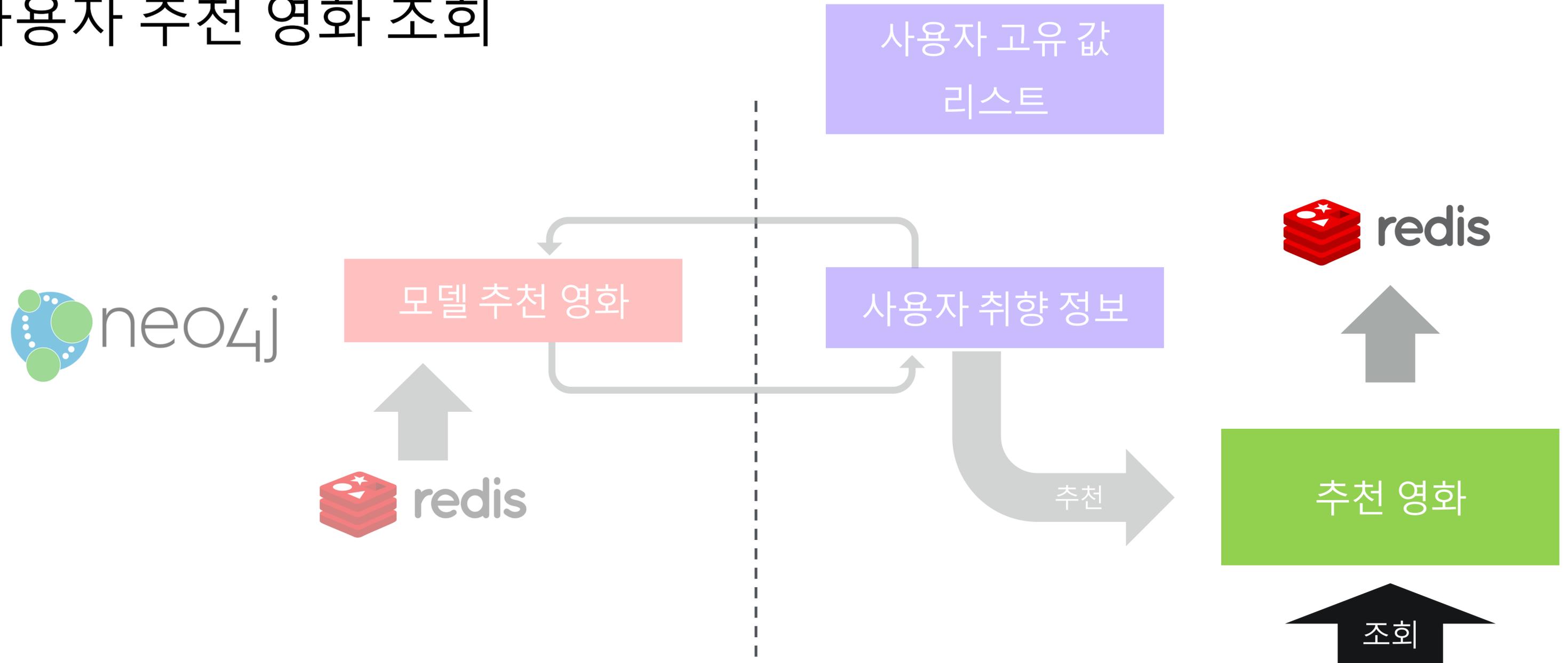
5.3 상시 업데이트

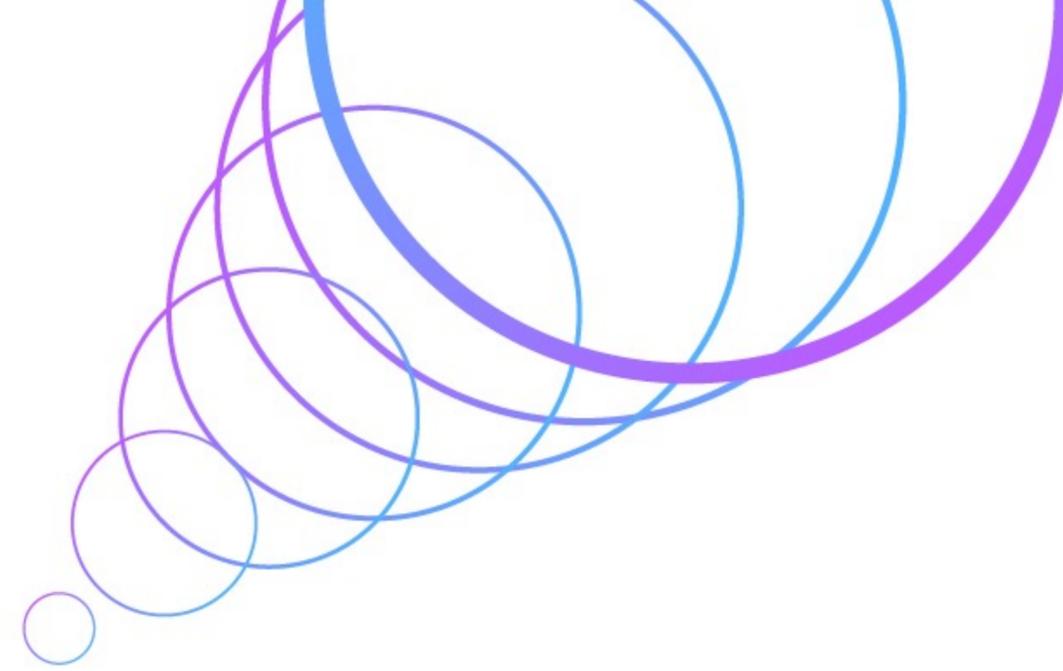
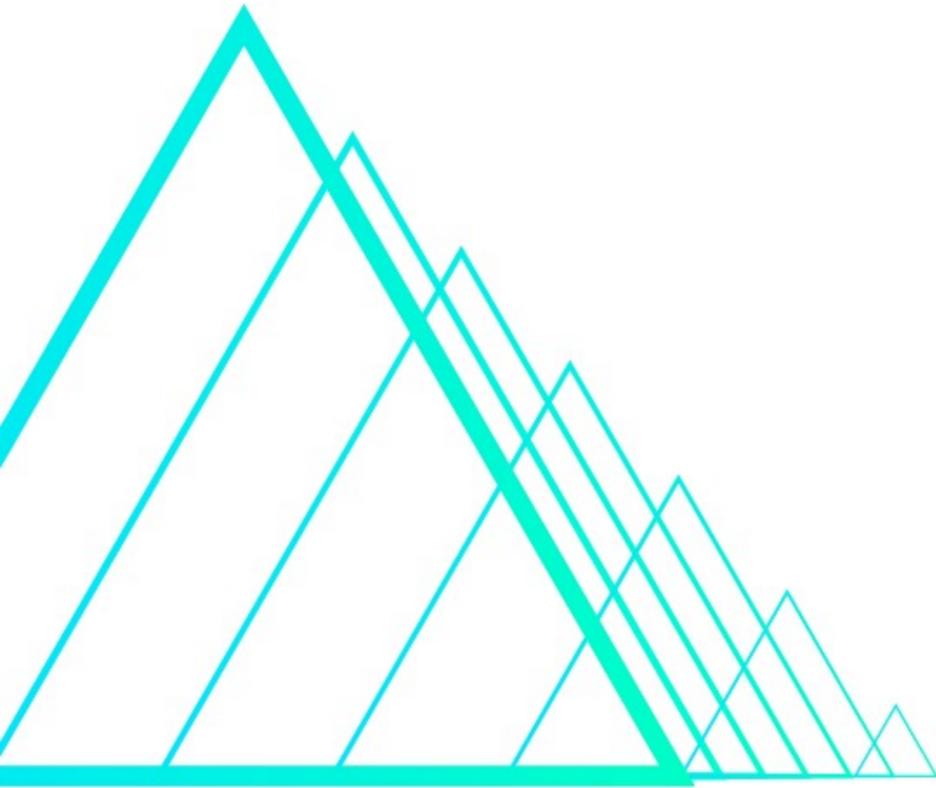
사용자 취향 별 추천



5.4 조회

사용자 추천 영화 조회





Thank You

