Special Topic: Recommender Systems

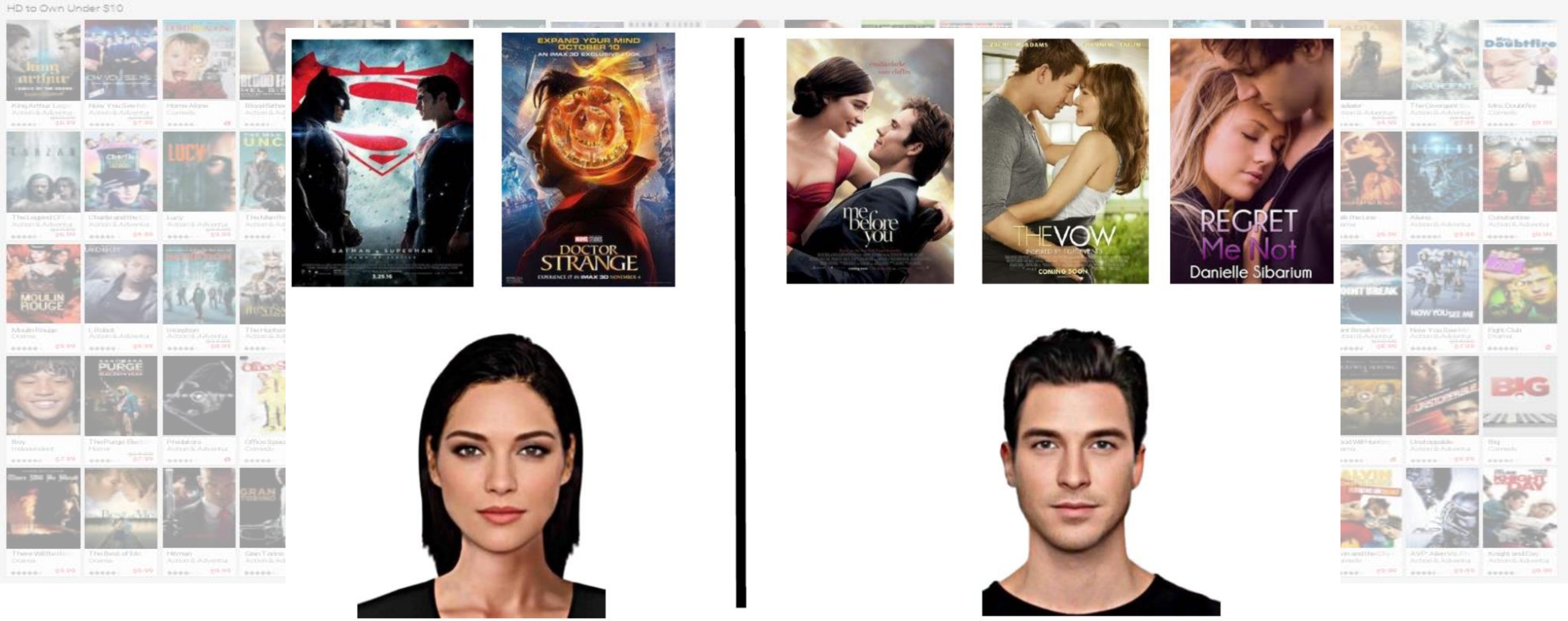
Dong-Kyu Chae

PI of the Data Intelligence Lab @HYU
Department of Computer Science & Data Science
Hanyang University

Data
Intelligence
LAB

Recommender Systems (Personalization, Curation)

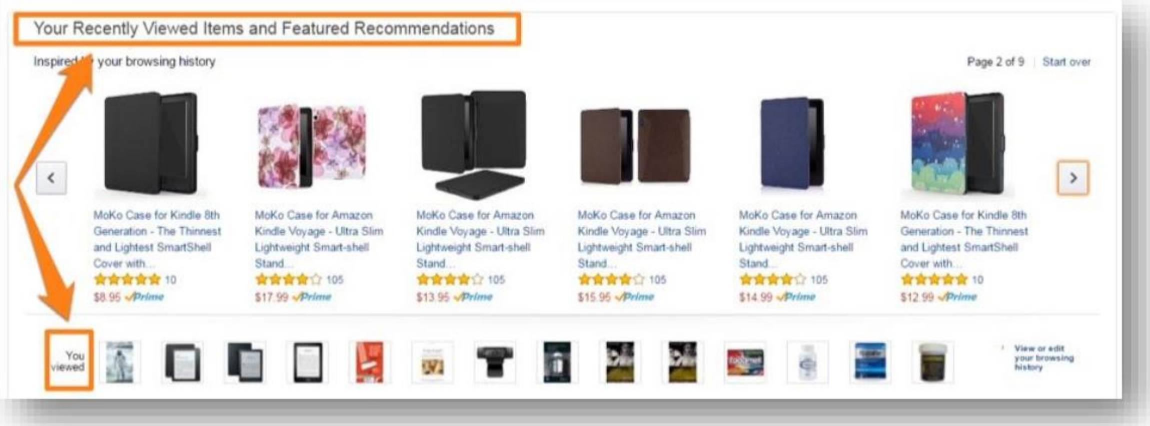
- Too many items, limited time
- Recommender system analyzes preferences of each user, and provides personalized top-N items with users



NETFLIX Flixster IMDb

Recommender Systems

❑ Amazon



Your Recently Viewed Items and Featured Recommendations

Inspired by your browsing history

Page 2 of 9 | Start over

MoKo Case for Kindle 8th Generation - The Thinnest and Lightest SmartShell Cover with... 10
$8.95 ✓Prime

MoKo Case for Amazon Kindle Voyage - Ultra Slim Lightweight Smart-shell Stand... 105
$17.99 ✓Prime

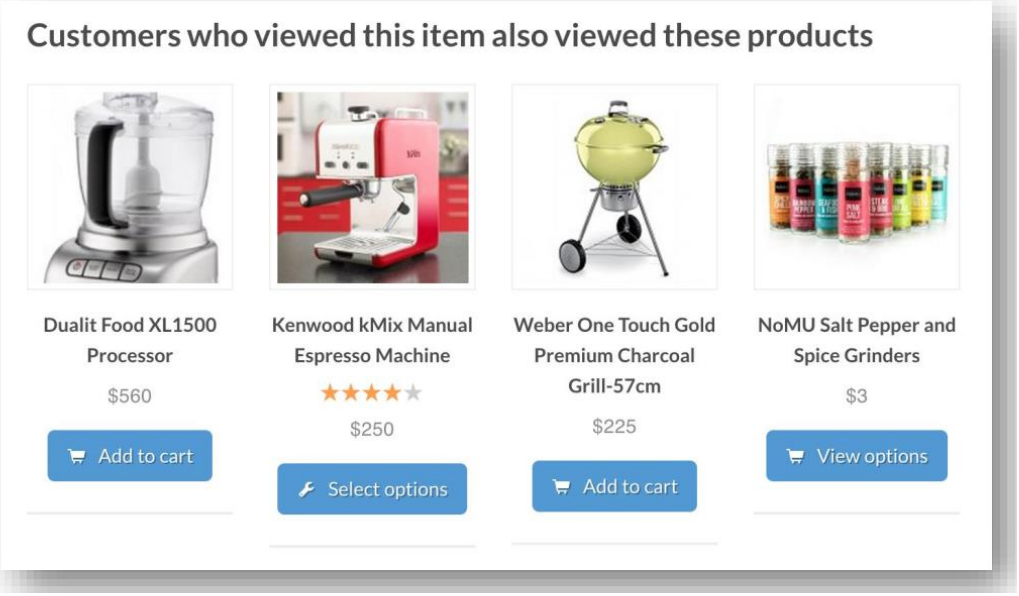
MoKo Case for Amazon Kindle Voyage - Ultra Slim Lightweight Smart-shell Stand... 105
$13.95 ✓Prime

MoKo Case for Amazon Kindle Voyage - Ultra Slim Lightweight Smart-shell Stand... 105
$15.95 ✓Prime

MoKo Case for Amazon Kindle Voyage - Ultra Slim Lightweight Smart-shell Stand... 105
$14.99 ✓Prime

MoKo Case for Kindle 8th Generation - The Thinnest and Lightest SmartShell Cover with... 10
$12.99 ✓Prime

Amazon product recommendation



Customers who viewed this item also viewed these products

Dualit Food XL1500 Processor
$560

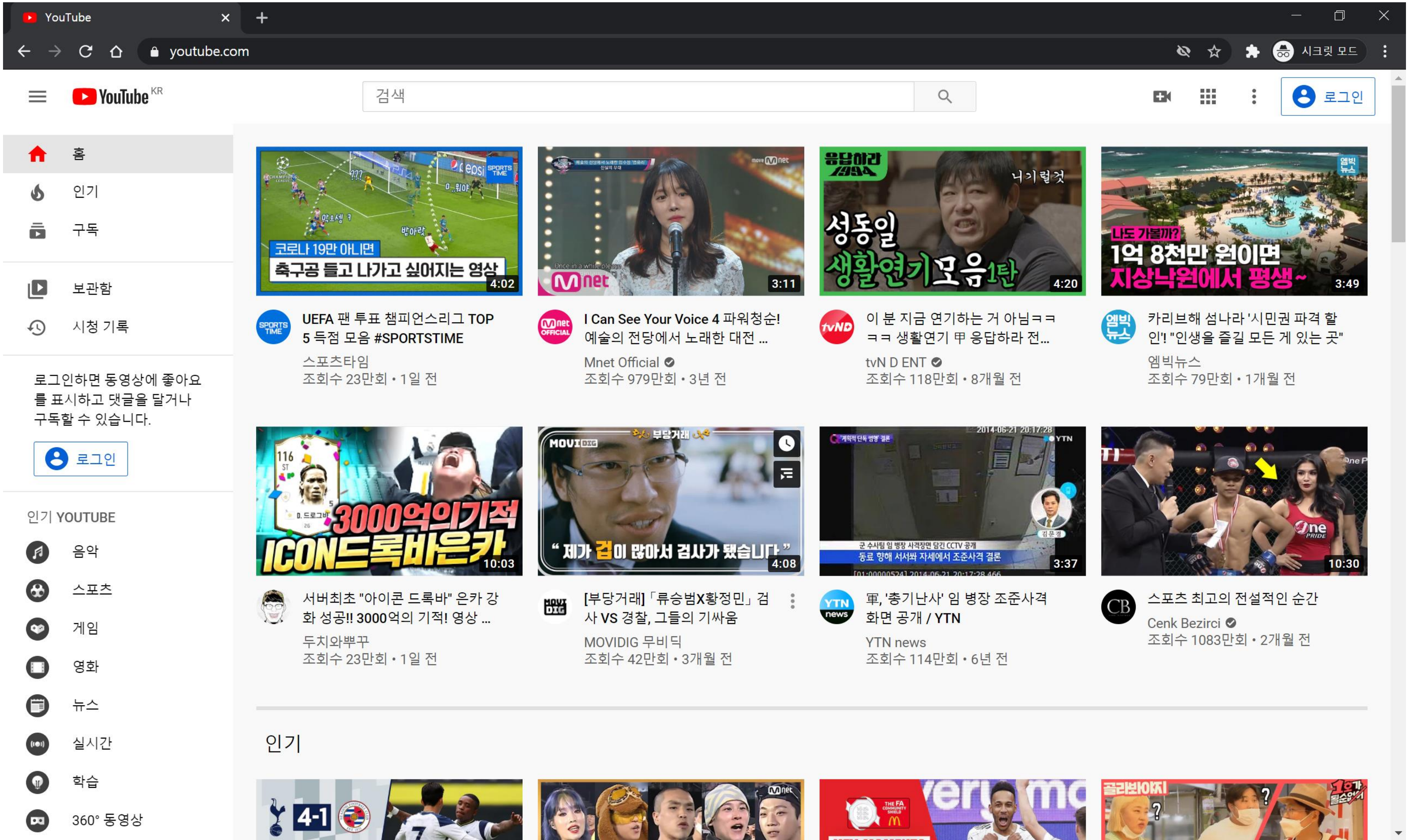
Kenwood kMix Manual Espresso Machine
★★★★☆
$250

Weber One Touch Gold Premium Charcoal Grill-57cm
$225

NoMU Salt Pepper and Spice Grinders
$3

Recommender Systems

❑ Youtube



로그인하면 동영상에 좋아요를 표시하고 댓글을 달거나 구독할 수 있습니다.

UEFA 팬 투표 챔피언스리그 TOP 5 득점 모음 #SPORTSTIME
스포츠타임
조회수 23만회 • 1일 전

I Can See Your Voice 4 파워청순! 예술의 전당에서 노래한 대전 ...
Mnet Official
조회수 979만회 • 3년 전

이 분 지금 연기하는 거 아님ㅋㅋㅋㅋ 생활연기 甲 응답하라 전...
tvN D ENT
조회수 118만회 • 8개월 전

카리브해 섬나라 '시민권 파격 할인'! "인생을 즐길 모든 게 있는 곳"
엠빅뉴스
조회수 79만회 • 1개월 전

서버최초 "아이콘 드록바" 은카 강화 성공!! 3000억의 기적! 영상 ...
두치와뿌꾸
조회수 23만회 • 1일 전

[부당거래]「류승범X황정민」검사 VS 경찰, 그들의 기싸움
MOVIDIG 무비딕
조회수 42만회 • 3개월 전

軍, '총기난사' 임 병장 조준사격 화면 공개 / YTN
YTN news
조회수 114만회 • 6년 전

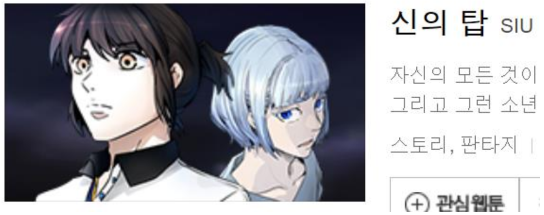
스포츠 최고의 전설적인 순간
Cenk Bezirci
조회수 1083만회 • 2개월 전

Recommender Systems

☐ Netflix

Recommender Systems

- ❑ **Cartoon**

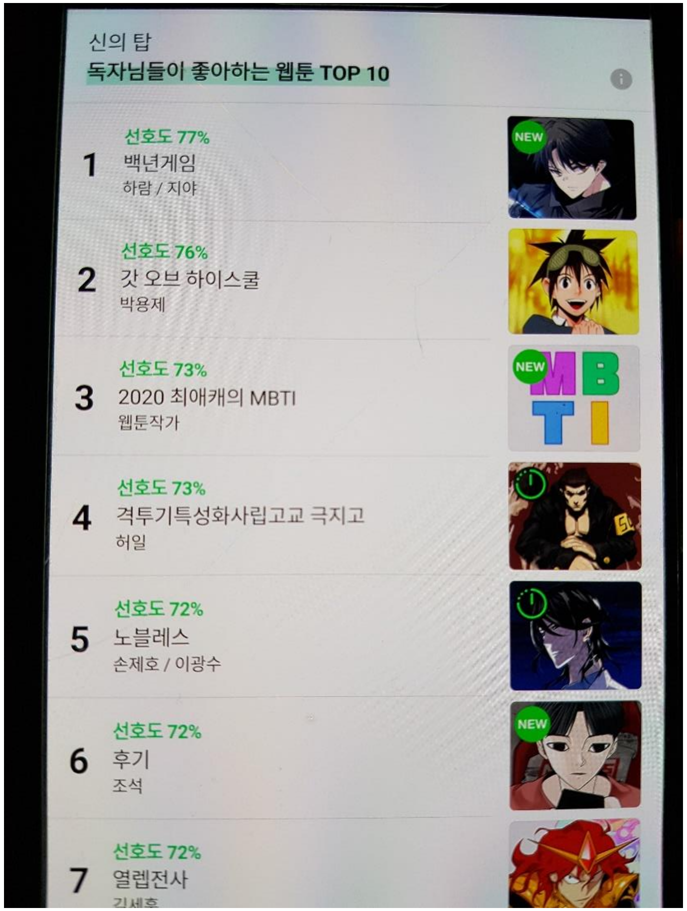


신의 탑 SIU

자신의 모든 것이
그리고 그런 소년

스토리, 판타지 |

⊕ 관심웹툰



신의 탑
독자님들이 좋아하는 웹툰 TOP 10

1 선호도 77%
백년게임
하람 / 지야

2 선호도 76%
갓 오브 하이스쿨
박용제

3 선호도 73%
2020 최애캐의 MBTI
웹툰작가

4 선호도 73%
격투기특성화사립고교 극지고
허일

5 선호도 72%
노블레스
손제호 / 이광수

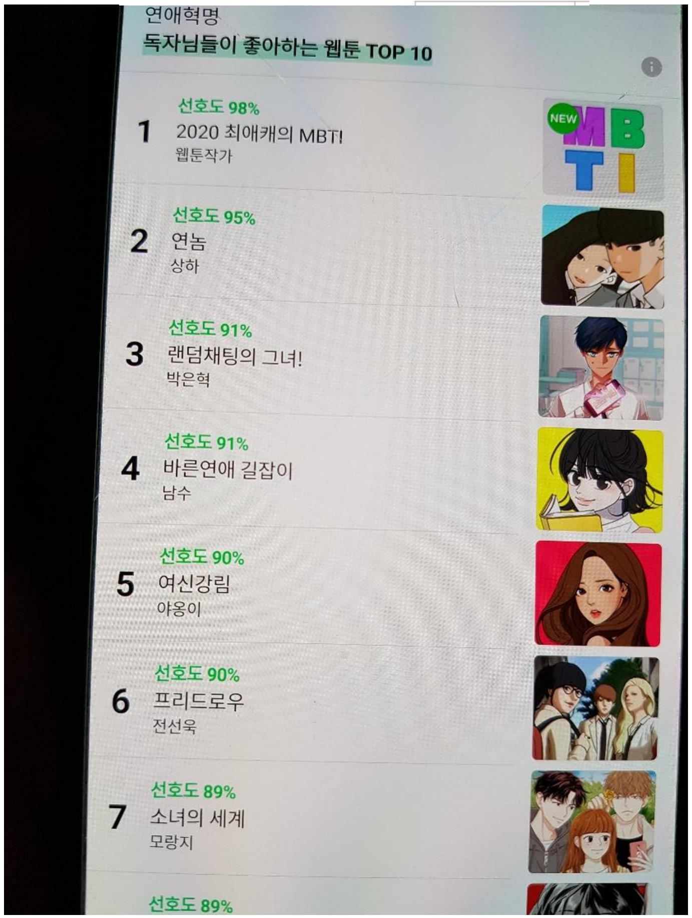
6 선호도 72%
후기
조석

7 선호도 72%
열렙전사

연애혁명 2

로맨스, 그런 건
신개념 개그 로

스토리, 드라마



연애혁명
독자님들이 좋아하는 웹툰 TOP 10

1 선호도 98%
2020 최애캐의 MBTI
웹툰작가

2 선호도 95%
연놈
상하

3 선호도 91%
랜덤채팅의 그녀!
박은혁

4 선호도 91%
바른연애 길잡이
남수

5 선호도 90%
여신강림
야옹이

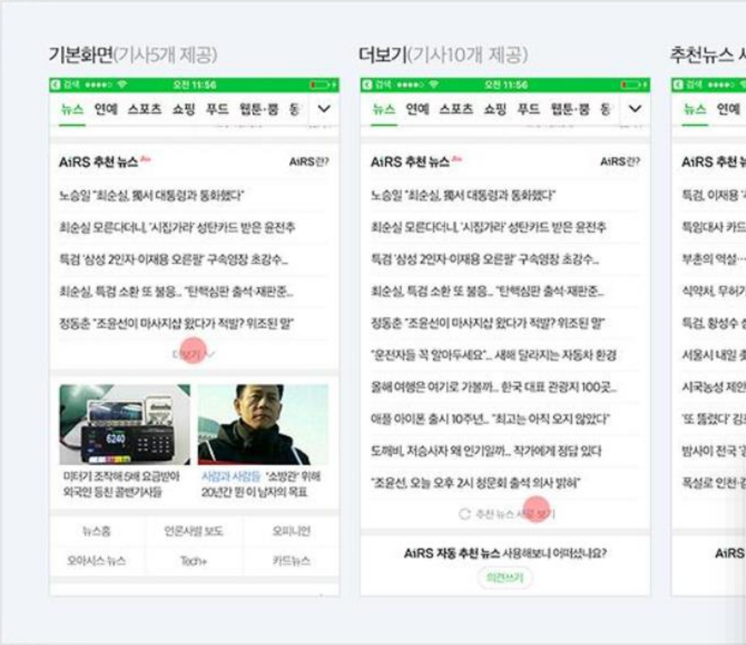
6 선호도 90%
프리드로우
전선욱

7 선호도 89%
소녀의 세계
모랑지

선호도 89%

Recommender Systems

❑ News recommendation



기본화면(기사5개 제공)

뉴스 연예 스포츠 쇼핑 푸드 웹툰·툰 등

AiRS 추천 뉴스 AiRS란?

노승일 "최순실, 獨서 대통령과 통화했다"

최순실 모른다더니, '시집가라' 성탄카드 받은 윤전추

특검 '삼성 2인자 이재용 오른팔' 구속영장 초강수..

최순실, 특검 소환 또 불응.. "탄핵심판 출석·재판준..

정동춘 "조윤선이 마사지샵 왔다가 적발? 위조된 말"

더보기

미터기 조작해 5배 요금받아 외국인 등친 콜밴기사들

사람과 사람들 '소방관' 위해 20년간 뛴 이 남자의 목표

뉴스홈 | 언론사별 보도 | 오피니언

오아시스 뉴스 | Tech+ | 카드뉴스

더보기(기사10개 제공)

뉴스 연예 스포츠 쇼핑 푸드 웹툰·툰 등

AiRS 추천 뉴스 AiRS란?

노승일 "최순실, 獨서 대통령과 통화했다"

최순실 모른다더니, '시집가라' 성탄카드 받은 윤전추

특검 '삼성 2인자 이재용 오른팔' 구속영장 초강수..

최순실, 특검 소환 또 불응.. "탄핵심판 출석·재판준..

정동춘 "조윤선이 마사지샵 왔다가 적발? 위조된 말"

"운전자들 꼭 알아두세요"... 새해 달라지는 자동차 환경

올해 여행은 여기로 가볼까.. 한국 대표 관광지 100곳..

애플 아이폰 출시 10주년.. "최고는 아직 오지 않았다"

도깨비, 저승사자 왜 인기일까.. 작가에게 정답 있다

"조윤선, 오늘 오후 2시 청문회 출석 의사 밝혀"

추천 뉴스 새로 보기

AiRS 자동 추천 뉴스 사용해보니 어떠셨나요?

의견쓰기

추천뉴스 새로보기(기사10개 새로추천)

뉴스 연예 스포츠 쇼핑 푸드 웹툰·툰 등

AiRS 추천 뉴스 AiRS란?

특검, 이재용 '추가수사' 증거수집.. 황성수 전무 전격소환

특임대사 카드로 트럼프 행정부와 관계 물꼬 튼다

부촌의 역설…서울내 빈집 최다지역은 강남

식약처, 무허가 원료로 만든 홍삼제품 5종 회수명령

특검, 황성수 삼성전자 전무 소환 조사 중

서울시 내일 촛불집회 안전요원 238명 투입

시국농성 제안 법률가들 "법원 이재용 구속영장 발부하라"

'또 뚫렸다' 김포서 AI 세 번째 발생.. 도살처분·방역강화

밤사이 전국

폭설로 인천·

AiRS

Naver news recommendation

Facebook news recommendation

Cyborg Menace

Humanity hangs on the brink of extinction. New levels just added. Are you game?

Lstn

Tap into a world of fresh new audio. Listen to our newest playlist, Back to Fall Beats.

J. Somerfields

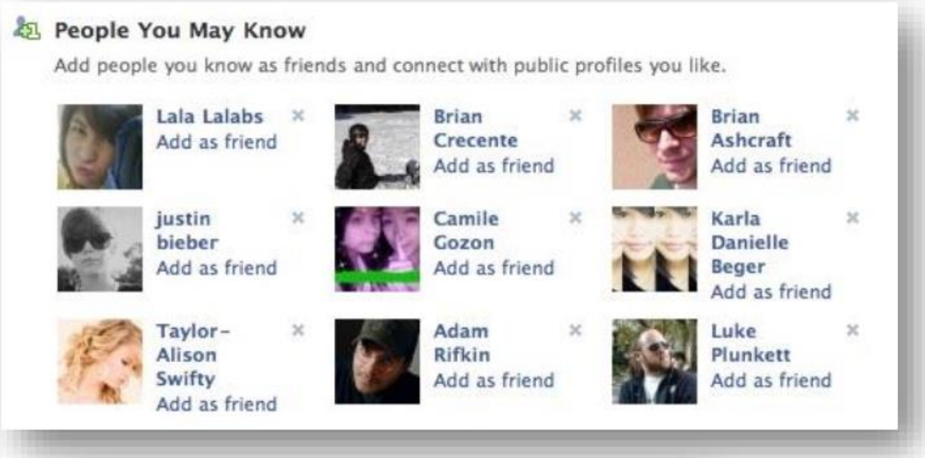
With the J. Somerfields App, the latest looks are always at your fingertips.

Nicole Gregory commented on this.

Many Applications

- Etc...

Social recommendation

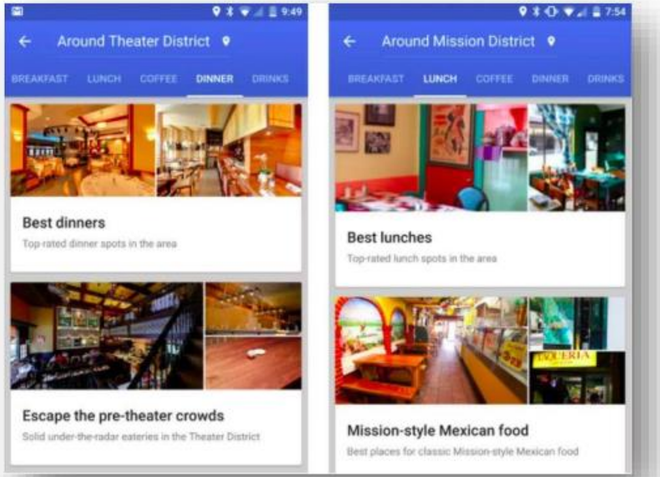


People You May Know

Add people you know as friends and connect with public profiles you like.

Lala Lalabs – Add as friend
Brian Crecente – Add as friend
Brian Ashcraft – Add as friend
justin bieber – Add as friend
Camile Gozon – Add as friend
Karla Danielle Beger – Add as friend
Taylor-Alison Swifty – Add as friend
Adam Rifkin – Add as friend
Luke Plunkett – Add as friend

Restaurant recommendation



Best dinners
Top-rated dinner spots in the area

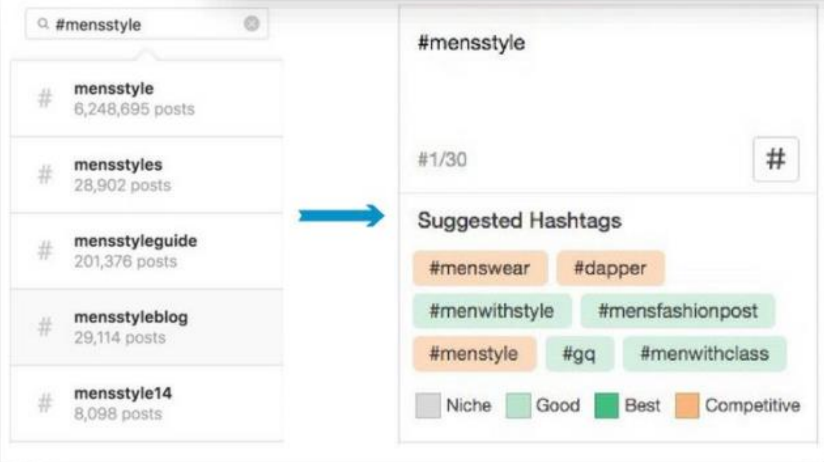
Escape the pre-theater crowds
Solid under-the-radar eateries in the Theater District

Best lunches
Top-rated lunch spots in the area

Mission-style Mexican food
Best places for classic Mission-style Mexican food

All-in-one Camera ★★★★★ (6) $1.79
The Night Sky ★★★★★ (38) $0.99
SCRABBLE (Kindle Tablet... ★★★★★ (1671) $2.99
Weather+ ★★★★★ (259) $0.99

Bad Piggies Premium HD... ★★★★★ (193) $2.99
A Pumpkin Maker ★★★★★ (13) $0.99
Rayman Jungle Run ★★★★★ (27) $2.99
WorldCraft - creative block... ★★★★★ (47) $0.99



#mensstyle

mensstyle 6,248,695 posts
mensstyles 28,902 posts
mensstyleguide 201,376 posts
mensstyleblog 29,114 posts
mensstyle14 8,098 posts

#mensstyle

#1/30

Suggested Hashtags

#menswear #dapper #menwithstyle #mensfashionpost #menstyle #gq #menwithclass

Niche Good Best Competitive

App recommendation

Tag recommendation

Many Applications

- Etc…

NETFLIX amazon NAVER kt

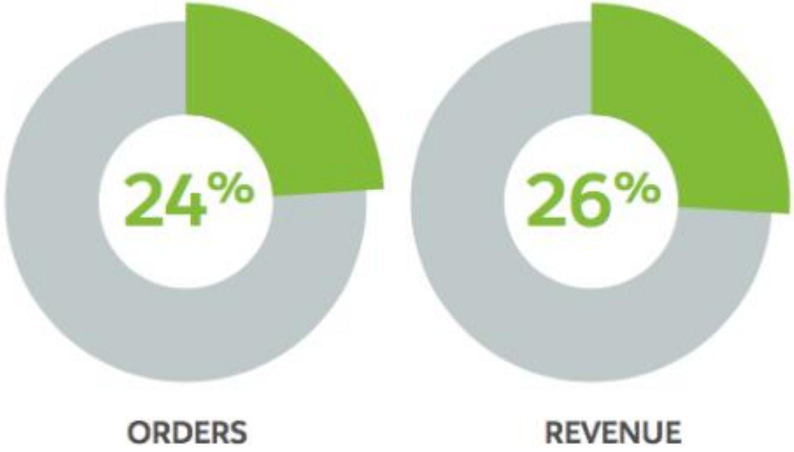
Google Instagram SK broadband kakao

YouTube facebook RIDIBOOKS WATCHA

Why Recommender Systems?

Importance of accurate recommendation

- **35% of the purchases** on **Amazon** are the result of their recommender systems
- **75% of watching** on **Netflix** comes from recommendations
- **70% of the traffics** are driven by “suggested videos” on **YouTube**.

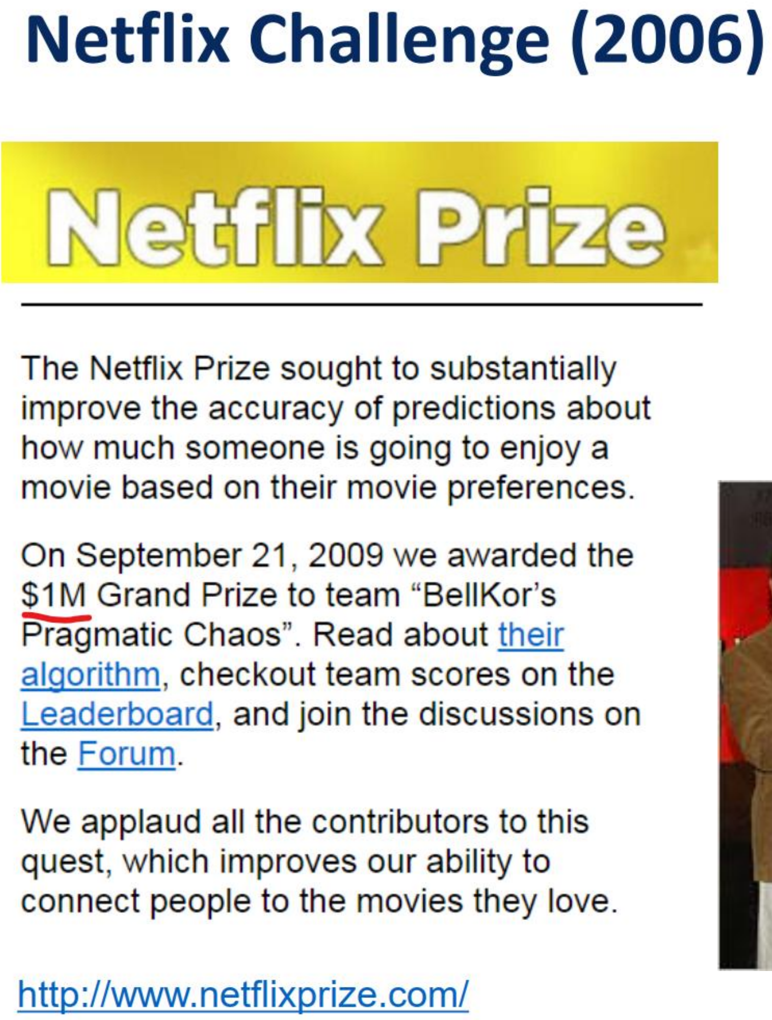


Contents

1. **Netflix Prize**
2. **Recommendation and Collaborative Filtering**
3. **KNN-based Methods**
4. **Matrix Factorization**
5. **Recent Recommenders**
6. **Case Study**

Netflix Prize

- **1 million dollar!**



Netflix Challenge (2006)

Netflix Prize

The Netflix Prize sought to substantially improve the accuracy of predictions about how much someone is going to enjoy a movie based on their movie preferences.

On September 21, 2009 we awarded the $1M Grand Prize to team "BellKor's Pragmatic Chaos". Read about their algorithm, checkout team scores on the Leaderboard, and join the discussions on the Forum.

We applaud all the contributors to this quest, which improves our ability to connect people to the movies they love.

http://www.netflixprize.com/

Competition

- $1 million
- Prize for 10% improvement on Netflix

Netflix Prize

❑ Problem setting

rating prediction task

- ❑ Training data

  open data

  - 100 million ratings, 480,000 users, 17,770 movies
  - 6 years of data: 2000-2005

- ❑ Test data

  close data

  - Last few ratings of each user (2.8 million)

- ❑ Evaluation criteria: **Root Mean Square Error (RMSE)**

$$\frac{1}{|R|}\sqrt{\sum_{(i,x)\in R}(\hat{r}_{xi}-r_{xi})^2}$$

→ lower is better

predict rating

actual rating

# of test data

- ❑ Netflix's system RMSE: 0.9514

  ↓ 0.85626

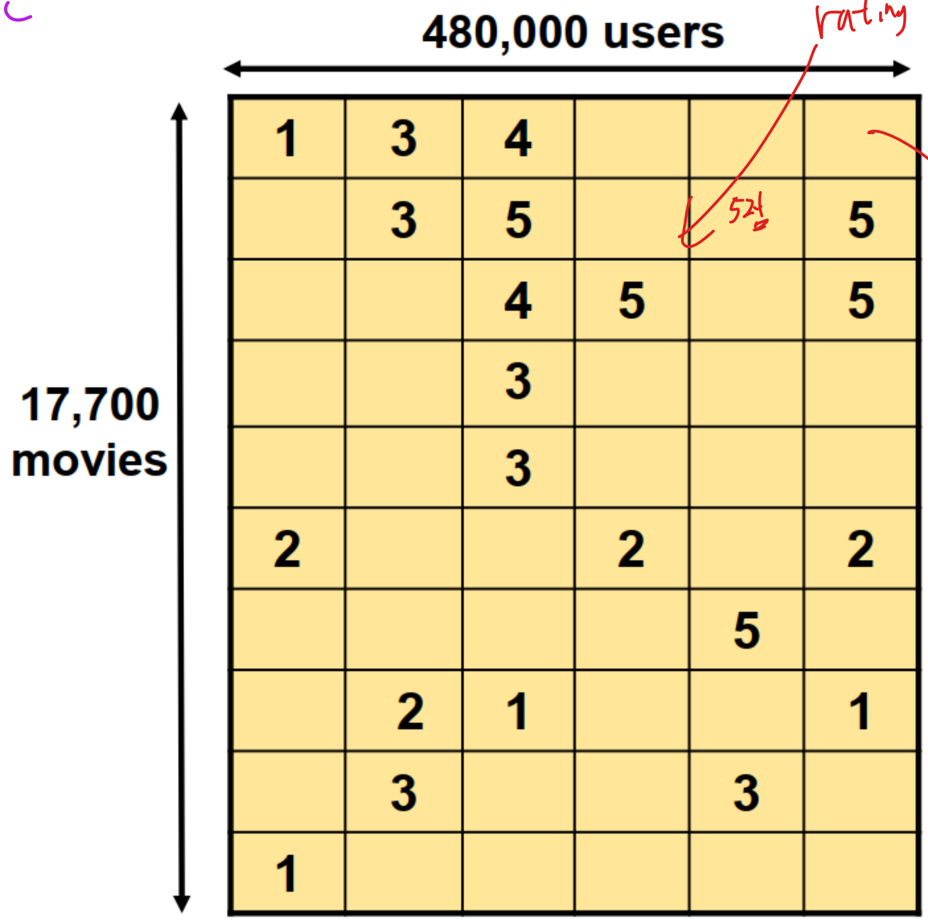
  10% improve

Netflix Prize

❑ User-Item Matrix R in Netflix

- ❑ Very sparse: less than 1% of entries are observed

빈칸



480,000 users

17,700 movies

1	3	4			
	3	5			5
		4	5		5
		3			
		3			
2			2		2
				5	
	2	1			1
	3			3	
1					

rating

5점

recommend system

R.S가

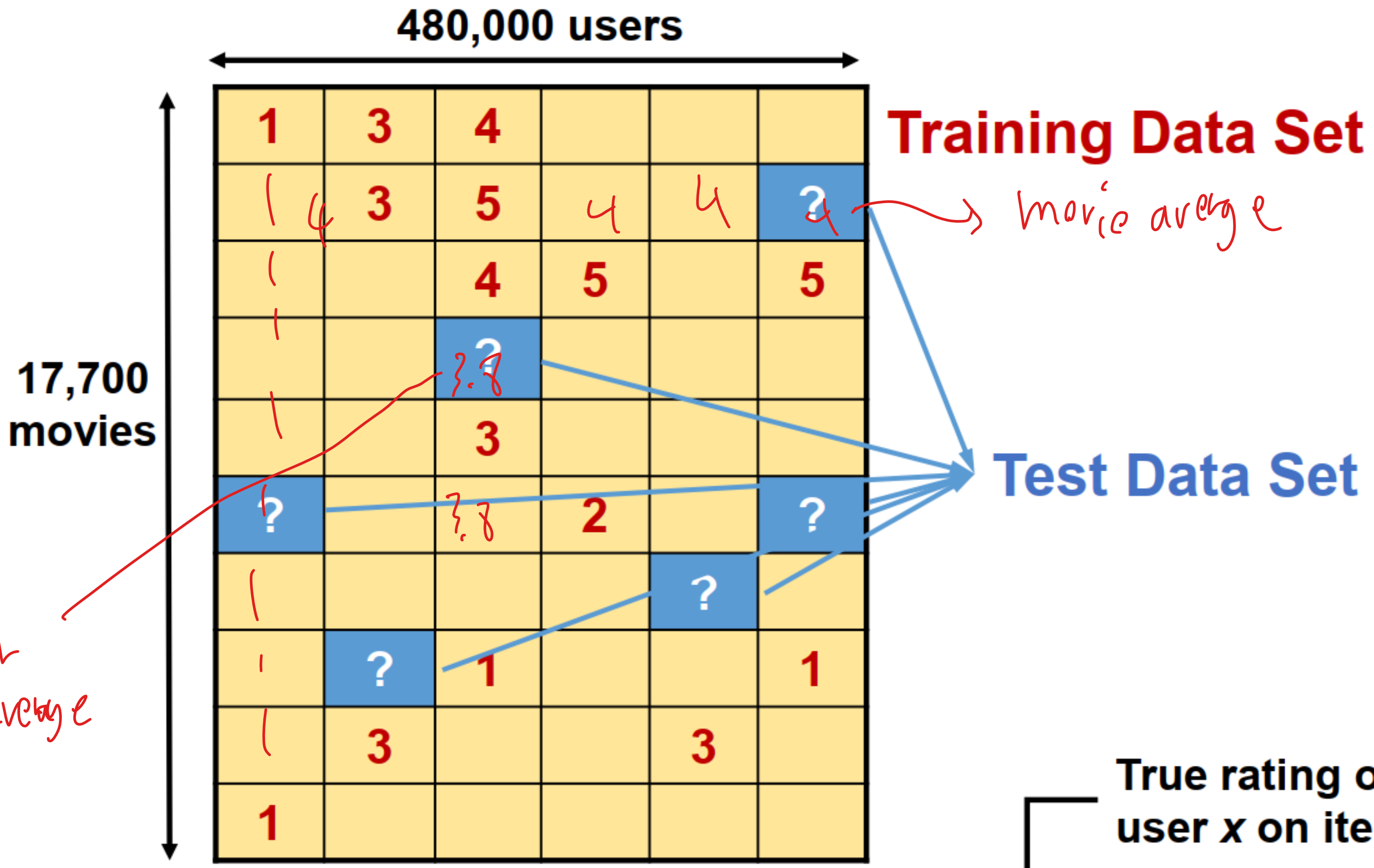
rating score를

예측함

6명이라가정

$\hat{r}_{61}$

Netflix Prize



1	3	4			
	4 3	5	4	4	? 4
		4	5		5
		? 3.8			
		3			
?		3.8	2		?
				?	
	?	1			1
	3			3	
1					

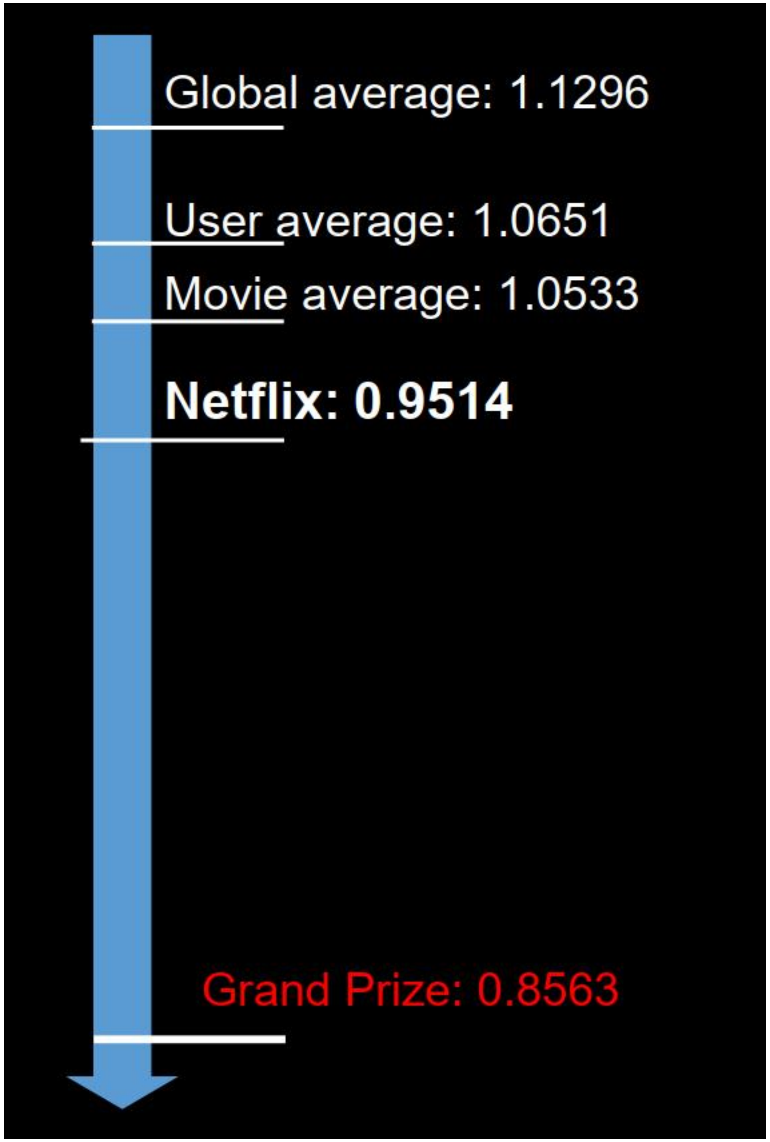
True rating of user *x* on item *i*

$$\textbf{RMSE} = \frac{1}{\underset{6}{|R|}}\sqrt{\sum_{(i,x)\in \underset{6}{R}}(\hat{r}_{xi} - r_{xi})^2}$$

Predicted rating

Netflix Prize

- **RMSE score overview**



Global average: 1.1296

User average: 1.0651

Movie average: 1.0533

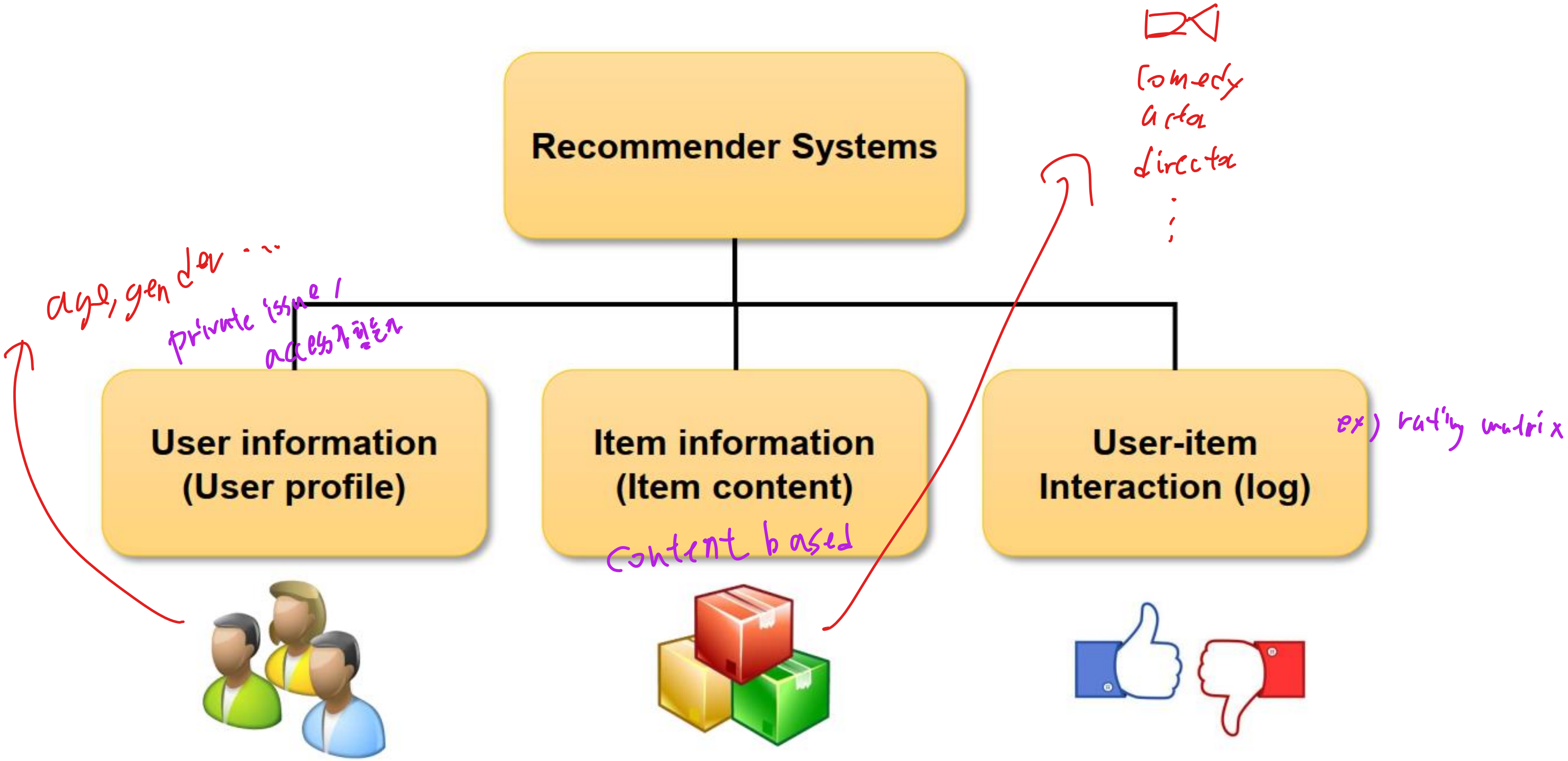
Netflix: 0.9514

Grand Prize: 0.8563

~ 2009

Category of Recommendation Algorithms

- ❑ **User profile matching**
- ❑ **Content-based recommendation**
- ❑ **Collaborative filtering (using user-item interactions)**



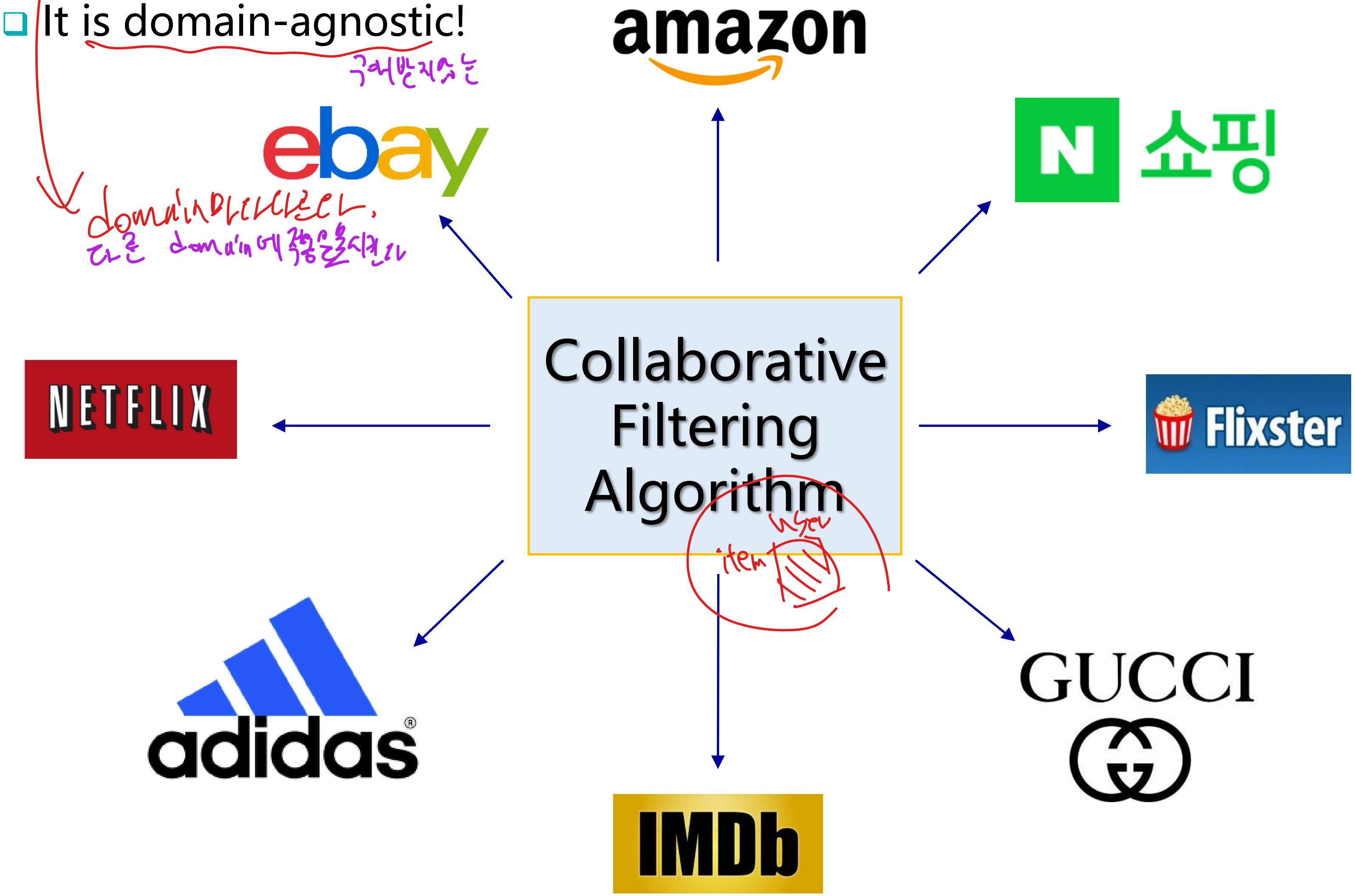
Contents-based VS Collaborative Filtering

일반화하다

- **Why collaborative filtering is more popular?**
  - It is domain-agnostic!

구애받지않는

domain마다다른데,
다른 domain에 적용을 시켜도



Collaborative Filtering Algorithm

Background

- Types of user-item matrix data
  - interaction
  - **Explicit feedback**: ratings, thumb up, like & dislike, etc... difficult to get
  - **Implicit feedback**: click, purchase, bookmark, etc... (popularized from 2008)

binary data easy to get

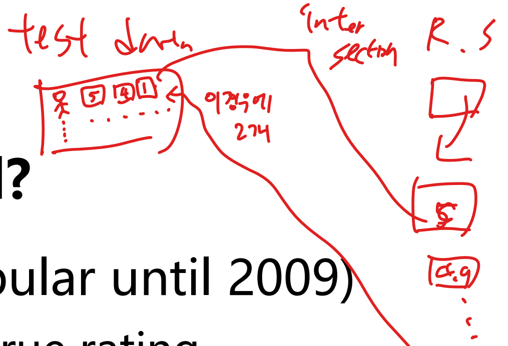
	Item					
User	1			3		
		3			2	
		4			4	
	4		3			1
		5			5	
			2			5

Rating matrix

	Item					
User		1			1	
			1			
	1			1		
			1			1
		1			1	
			1			1

Interaction matrix

Background



- ❑ **How to evaluate a recommendation method?**
  - ❑ Rating prediction perspective: **RMSE** (most popular until 2009)
    - How a predicted rating score is similar with the true rating

$$MAE = \frac{1}{N}\sum |predicted - actual|$$

or

$$RMSE = \sqrt{\frac{1}{N}\sum (predicted - actual)^2}$$

  - ❑ Classification perspective: **Precision, recall, F-measure (F1 score)**
    - How many a top-N recommendation list contains the true items – DZ "4" rating

$$P = \frac{\text{\# of our recommendations that are relevant}}{\text{\# of items we recommended}} \qquad r = \frac{\text{\# of our recommendations that are relevant}}{\text{\# of all the possible relevant items}}$$

  - ❑ Ranking perspective: **NDCG** (Normalized Discounted Cumulative Gain), **MRR** (Mean Reciprocal Rank)
    - How much the true items are highly ranked in the top-N recommendation list

Thank You