

추천시스템 구현 프로젝트

Recommender System

1. 추천시스템 이해

- 1) 추천시스템 개념
- 2) 추천시스템 방법론
- 3) 기업에서의 추천시스템
- 4) 추천시스템 Trend

2. 추천시스템 구현

- 1) 과거 추천시스템
- 2) 콘텐츠 기반의 추천시스템
- 3) 협업필터링 기반 추천시스템
- 4) 딥러닝 기반의 추천시스템

Recommender System 이해

■ 추천시스템 개념

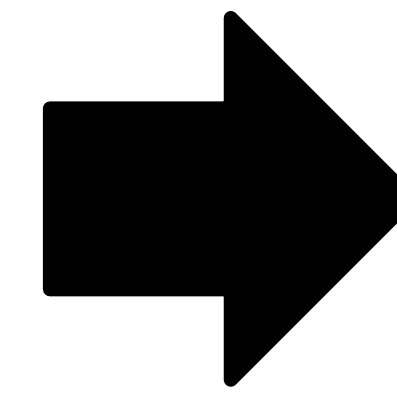
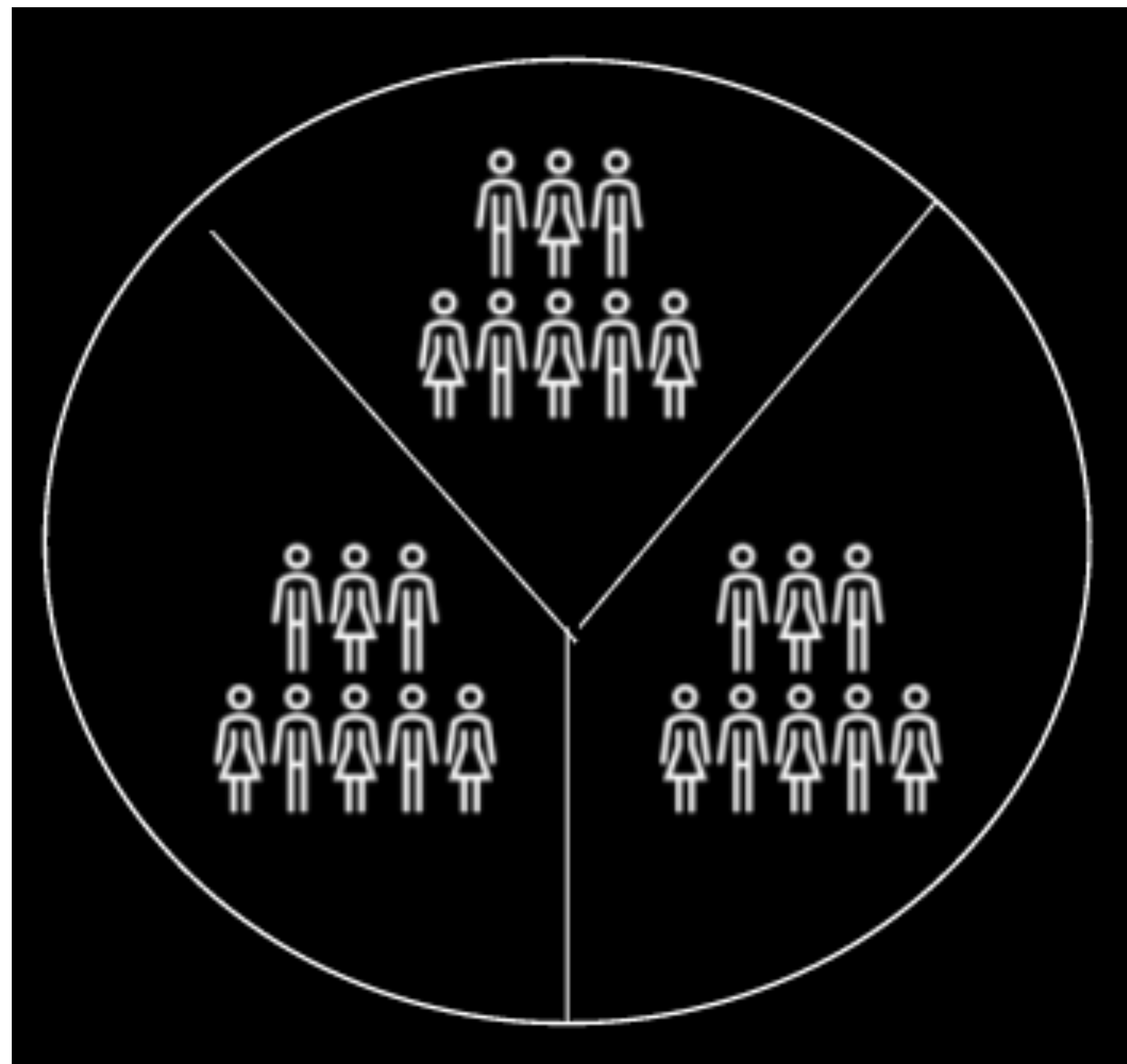
- 유저의 선호도 및 과거 행동을 바탕으로 **개인에 맞는 관심사를 제공**하는 분야
- 유튜브, 넷플릭스와 같은 **콘텐츠 추천**, 온라인 쇼핑몰이나 뉴스 추천, 금융 상품 추천, 검색 시스템 등 다양한 분야에서 사용
- 서비스 제공자 : quality가 높은 상품 제공으로 인한 **수익 창출**
- 유저 : 원하는 상품을 찾기 위한 **시간 절감**, 새로운 상품에 대한 **접근 용이**



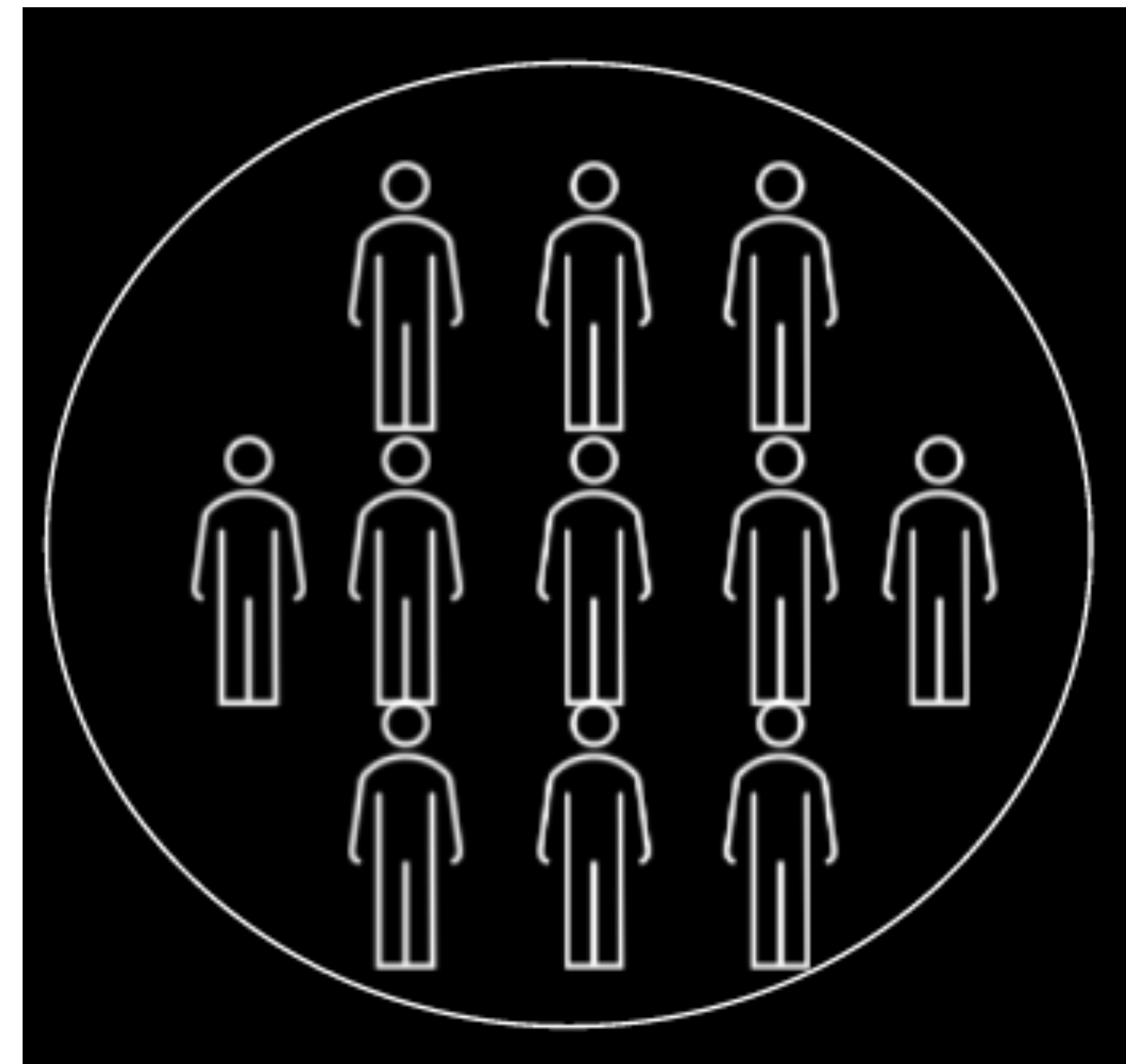
Recommender System 이해

- 추천시스템 개념

과거



현재

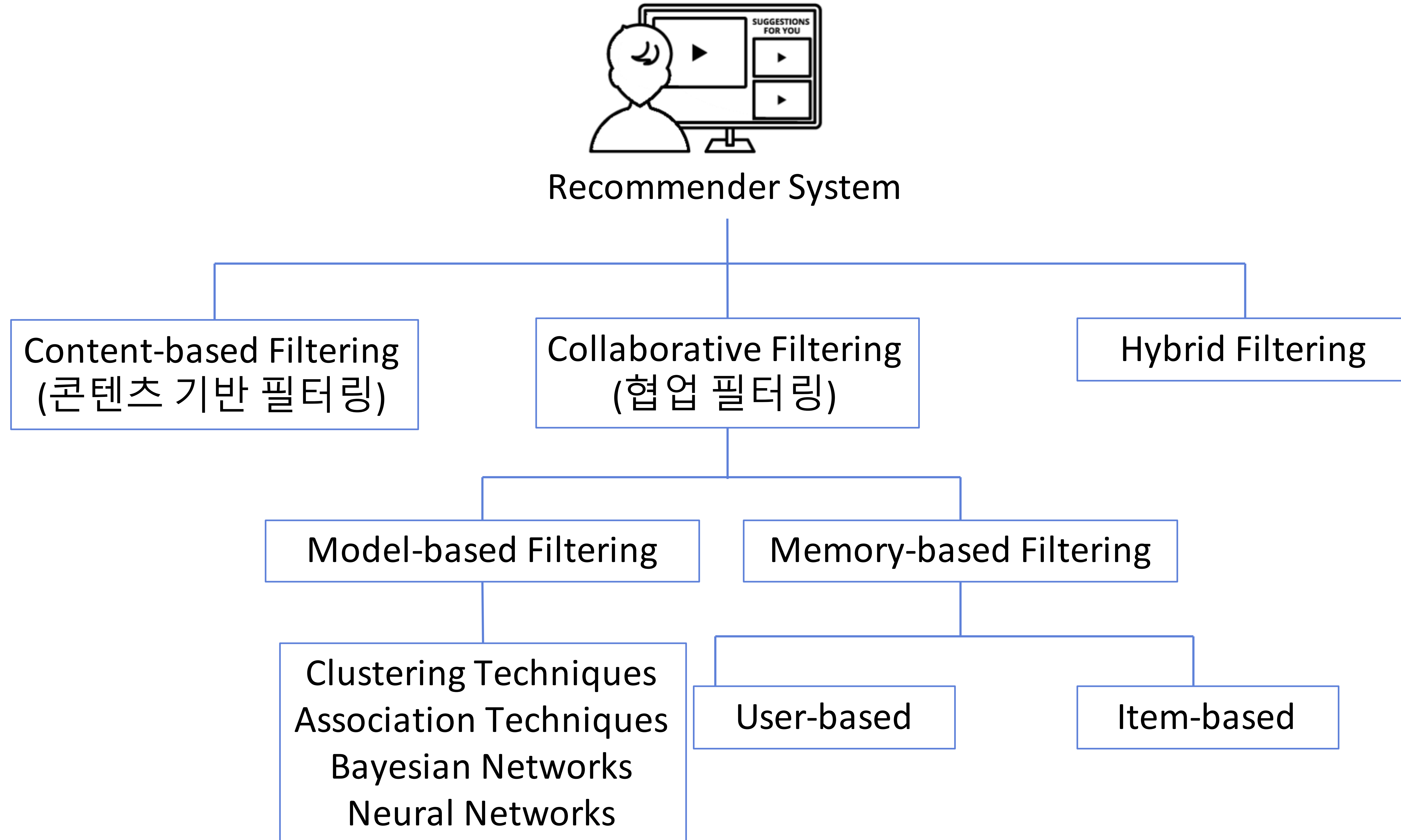


Recommender System 이해

- 추천시스템 방법론
 - 협업 필터링(Collaborative Filtering)
 - 내용 기반 필터링(Content-Based Filtering)
 - 지식 기반 필터링(Knowledge-Based Filtering)
 - 딥러닝(Deep Learning)
 - 하이브리드 필터링(협업필터링 & 딥러닝)

Recommender System 이해

■ 추천시스템 방법론



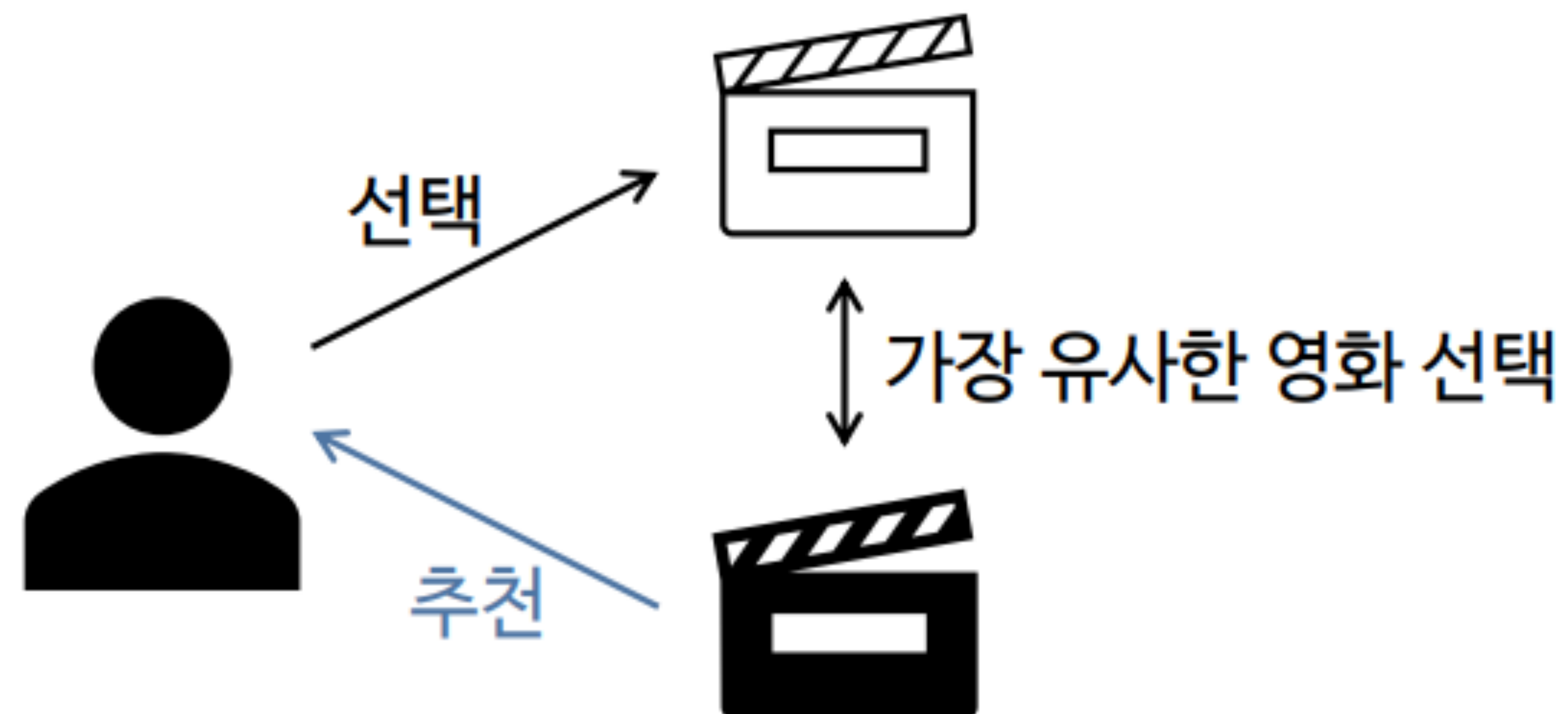
Recommender System 이해

■ Content-based Filtering

- 관심있는 아이템의 속성을 분석하여 새로운 아이템을 추천

(예) 내가 산 옷과 비슷하게 생긴 옷 추천, 관련된 뉴스 기사 추천 등

- 콘텐츠 기반으로 분석, 다른 유저의 정보가 사용되지 않음
- 사용자가 관심분야에 대해서 직접 입력한 정보나 구매 내역, 평점 등을 기반으로 선호 아이템 파악



Recommender System 이해

■ Content-based Filtering

- TF-IDF, Word2Vec 등의 Feature Extraction 방법론 사용
- 장점
 - 유저의 선호도에 대한 정보 없이, **아이템 정보만으로 추천 가능**
- 단점
 - 아이템을 설명할 수 있는 **메타데이터**(item's metadata) 구축이 필요
 - 이전에 구매 및 선택한 아이템과 비슷한 제품만 추천하는 **과도한 특수화** 발생

생

Recommender System 이해










- Collaborative Filtering (협업 필터링, CF)
 - User-Item Matrix로 부터 도출(User-Item interaction)되는 유사도 기반으로 아이템 추천
 - 메모리기반 협업 필터링
 - 추천을 위한 데이터를 메모리에 가지고 있으면서 이를 이용해서 추천하는 방식
 - User-Based, Item-Based
 - K Nearest Neighbors
 - 모델기반 협업 필터링
 - User-Item interaction을 머신러닝이나 딥러닝과 같은 모델을 이용하여 학습
 - 데이터로부터 추천을 위한 모델 구성 후 학습된 모델을 이용해서 추천
 - 장바구니 분석(Association rule)-아이템 간의 관계 학습, Clustering을 통해 유사한 아이템이
나

유저 간의 그룹 형성, Matrix Factorization, Bayesian Network, Decision Tree 등

Recommender System 이해

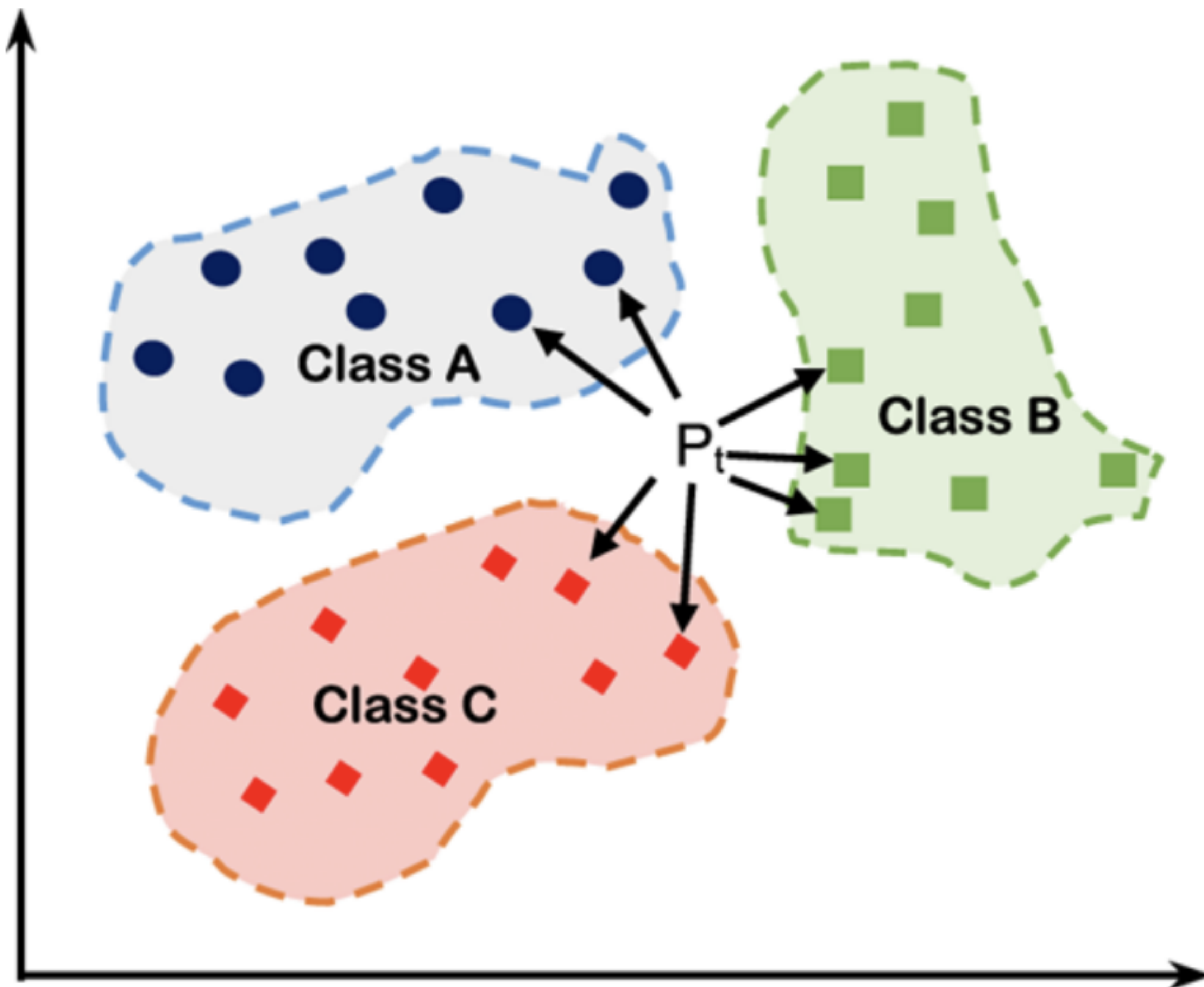
Collaborative Filtering (Memory-based Filtering)

- User-based : 사용자 간의 선호도(아이템에 대한 점수)나 구매 이력을 비교하여 추천하는 방법
- Item-based

		Item-based Similarity			
		 라면	 삼각김밥	 김치	 콜라
User-based Similarity	유저 A 	✓	✓	✓	✓
	유저 B 	✓	✓	✓	
	유저 C 	✓	✓		
	유저 D 	✓			
	유저 E 		✓		✓

Recommender System 이해

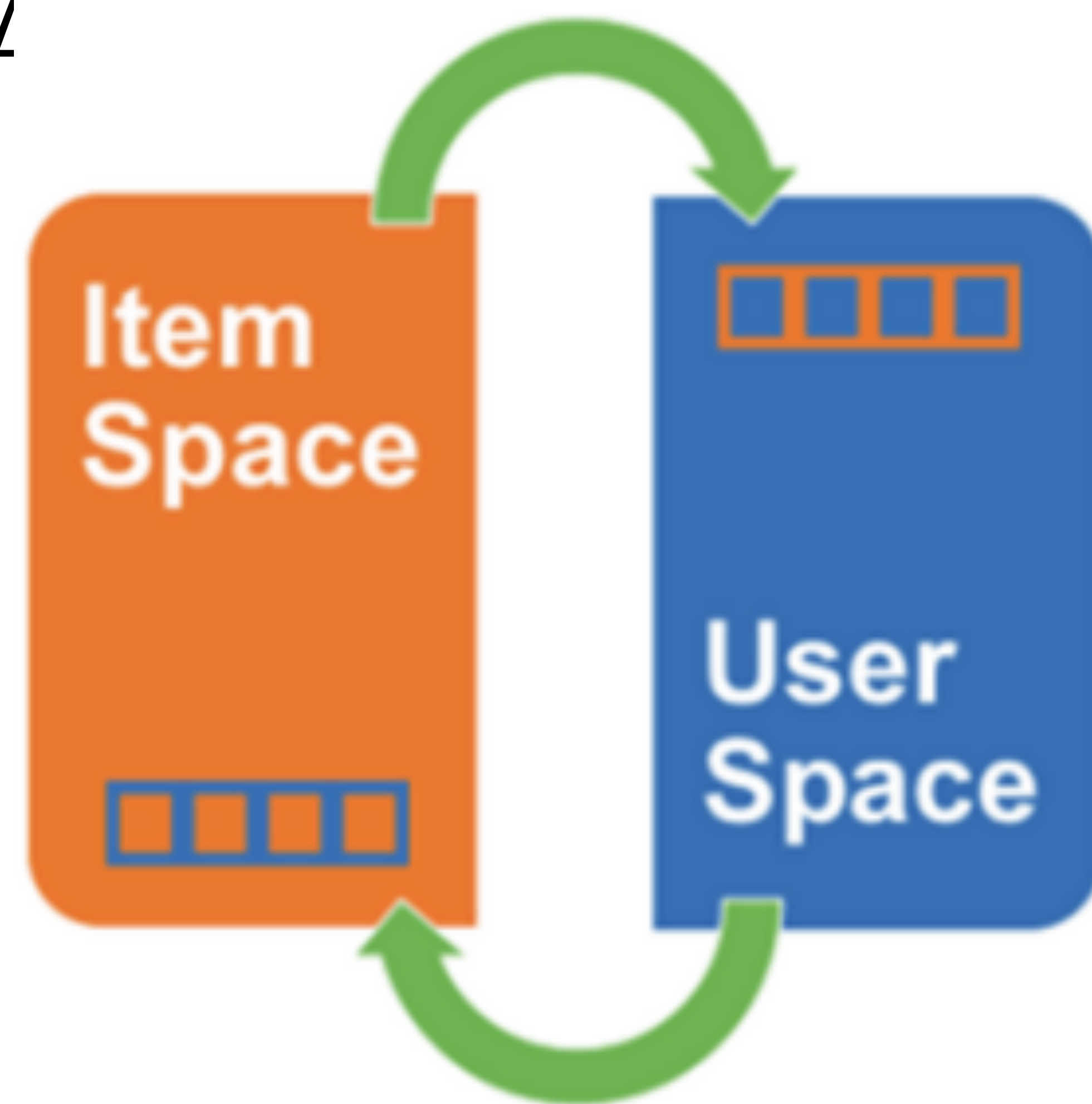
- Memory-based Filtering - K Nearest Neighbors
 - 가장 유사한 K 명의 Neighbors를 통해서 예측하는 방법



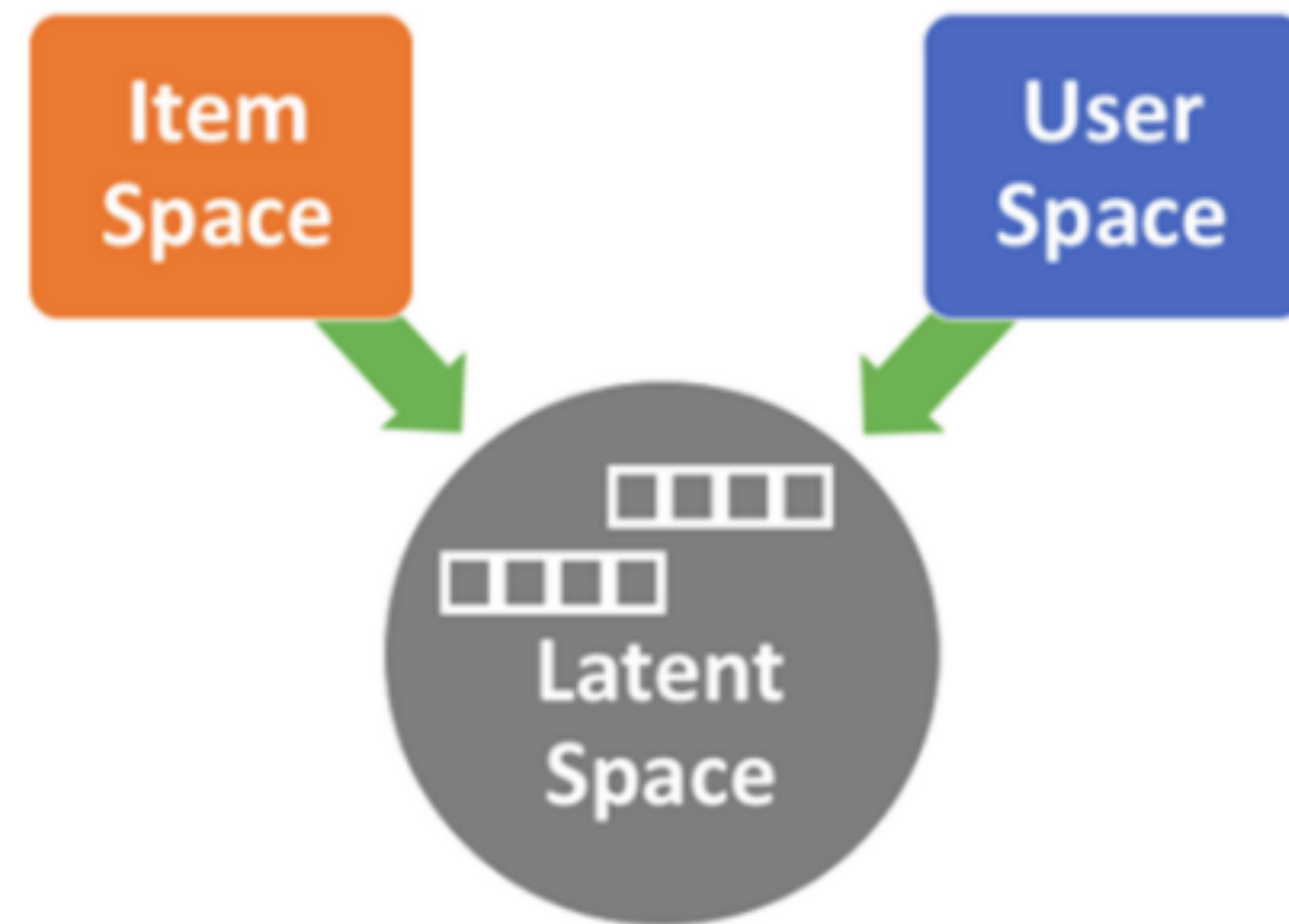
Recommender System 이해

■ Model-based Filtering - Matrix Factorization

- ‘사용자와 아이템 사이에는 사용자의 행동과 평점에 영향을 끼치는 잠재된 특성이 있을 것이다’
- User-Item Matrix를 F차원의 User와 Item의 latent factor 행렬곱으로 분해하는 방법
- SGD, /



Neighborhood Model

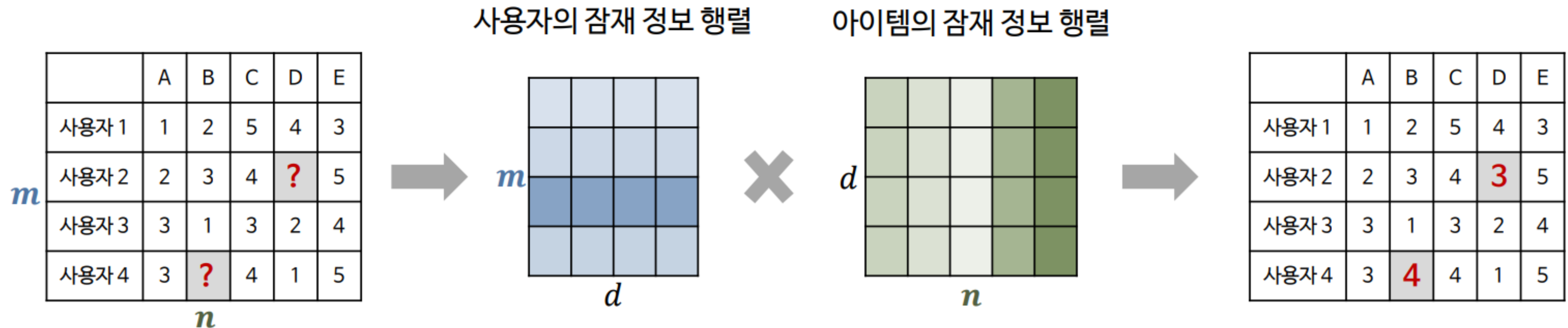


Latent Model

Recommender System 이해

■ Model-based Filtering - Matrix Factorization

- 잠재 요인 협업 필터링(Latent Factor Collaborative Filtering) 방식
- Rating Matrix에서 빈 공간을 채우기 위해서 사용자와 상품을 잘 표현하는 차원을 찾는 방법
- SVD, SGD, ALS 알고리즘(Latent Factor Collaborative Filtering) 사용



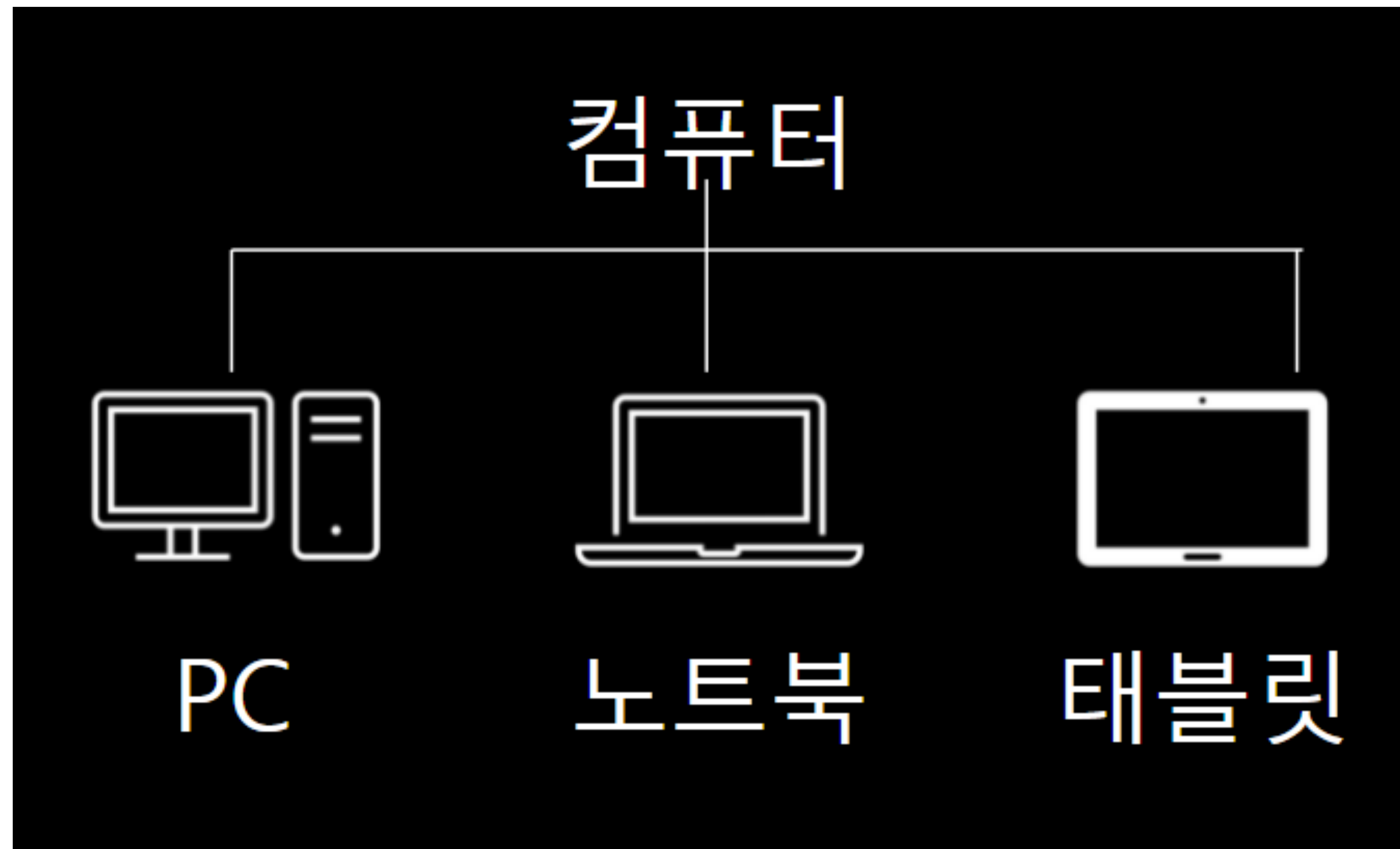
* 4명의 사용자가 5개의 아이템에 대한 평점을 매긴 경우

Recommender System 이해

- Collaborative Filtering (memory based & model based)
 - 장점 : 아이템에 대한 콘텐츠의 정보없이 사용 가능
 - 단점
 - 명시적 데이터를 구하기 힘들
 - Cold Start Problem : 새로운 유저나 아이템의 초기 정보가 부족한 경우 선호도 예측 불가능
 - Data Sparsity : 수많은 유저와 아이템 사이에 경험하지 못한, 구매해보지 못한 경우가 데이터의 대부분을 차지
 - Grey Sheep Problem : 일관성이 없는 의견을 가진 사용자들의 데이터는 추천에 혼란을 줌
 - Shilling Attack : 악의적으로 평가 점수를 긍정/부정으로 입력하는 경우 추천에 방해가 됨

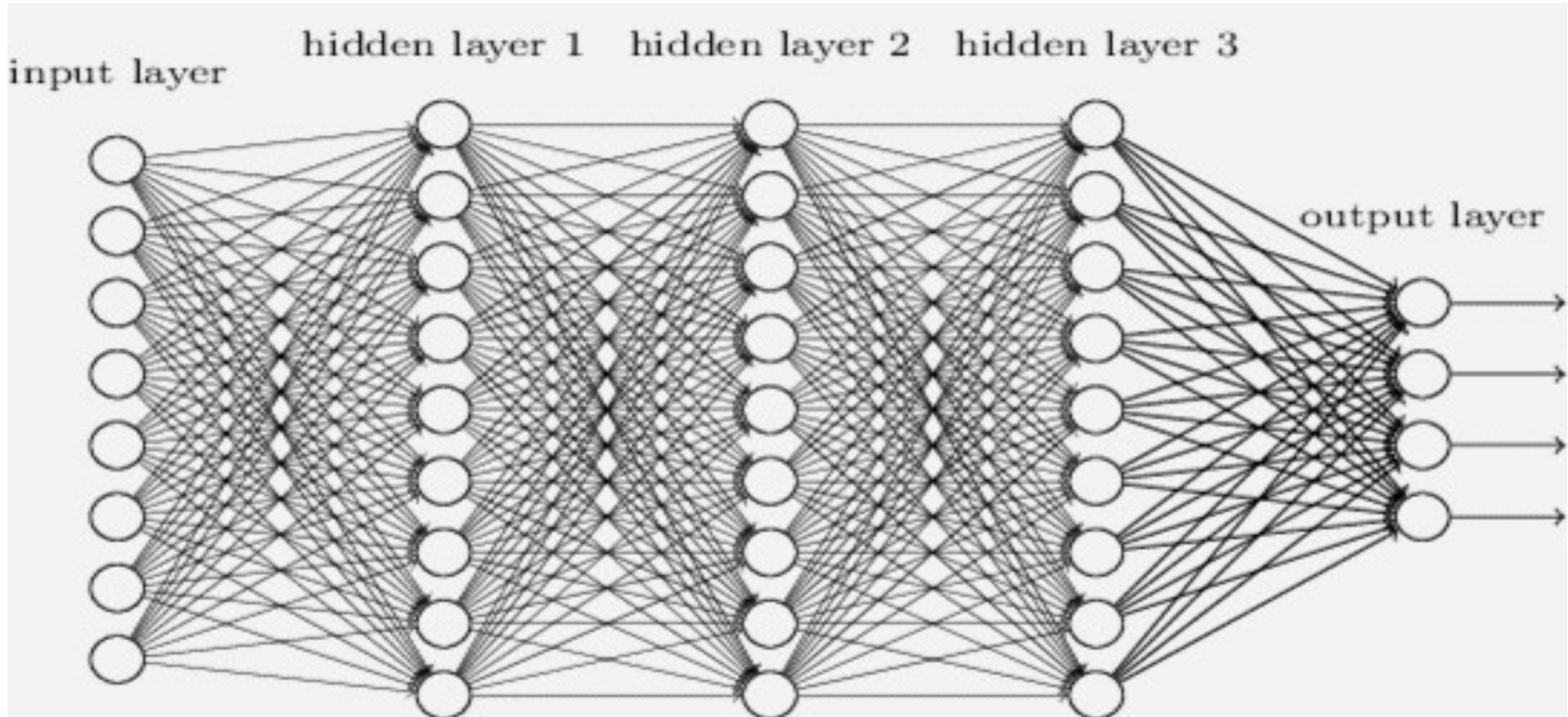
Recommender System 이해

- 지식기반 필터링(Knowledge-Based Filtering:KB)
 - 특정 분야 전문가의 도움을 받아서 그 분야에 대한 “전체적인 지식 구조”를 만들어서 활용하는 방법



Recommender System 이해

- 딥러닝(Deep Learning : DL) 추천 기술
 - 데이터 내에 존재하는 여러 추상적이며 복잡한 내용을 학습하여 핵심을 알아내는 딥러닝의 발전



Recommender System 이해

■ Neural Collaborative Filtering(논문)

- International World Wide Web Conference에서 2017년 발표된 논문
- 1708회 인용

Neural Collaborative Filtering

Xiangnan He
National University of
Singapore, Singapore
xiangnanhe@gmail.com

Liqiang Nie
Shandong University
China
nieliqiang@gmail.com

Lizi Liao
National University of
Singapore, Singapore
liaolizi.llz@gmail.com

Xia Hu
Texas A&M University
USA
hu@cse.tamu.edu

Hanwang Zhang
Columbia University
USA
hanwangzhang@gmail.com

Tat-Seng Chua
National University of
Singapore, Singapore
dcscts@nus.edu.sg

ABSTRACT

In recent years, deep neural networks have yielded immense success on speech recognition, computer vision and natural language processing. However, the exploration of deep neural networks on recommender systems has received relatively less scrutiny. In this work, we strive to develop techniques based on neural networks to tackle the key problem in recommendation — collaborative filtering — on the basis of implicit feedback.

Although some recent work has employed deep learning for recommendation, they primarily used it to model auxiliary information, such as textual descriptions of items and acoustic features of musics. When it comes to model the key factor in collaborative filtering — the interaction between user and item features, they still resorted to matrix factorization and applied an inner product on the latent features of users and items.

By replacing the inner product with a neural architecture that can learn an arbitrary function from data, we present a general framework named NCF, short for *Neural network-based Collaborative Filtering*. NCF is generic and can express and generalize matrix factorization under its framework. To supercharge NCF modelling with non-linearities, we propose to leverage a multi-layer perceptron to learn the user-item interaction function. Extensive experiments on two real-world datasets show significant improvements of our proposed NCF framework over the state-of-the-art methods. Empirical evidence shows that using deeper layers of neural networks offers better recommendation performance.

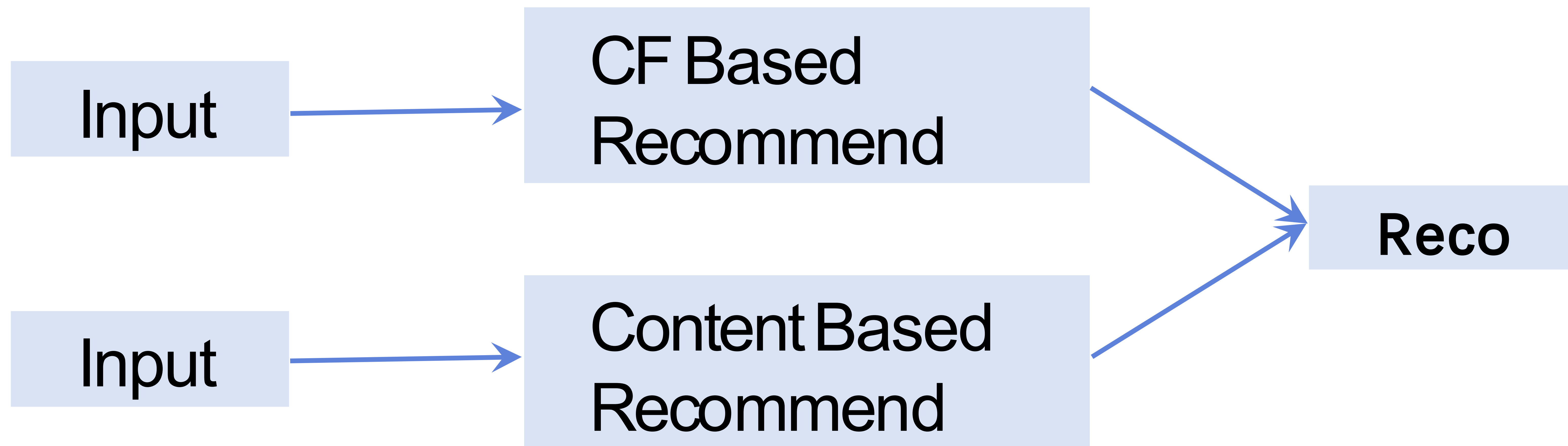
1. INTRODUCTION

In the era of information explosion, recommender systems play a pivotal role in alleviating information overload, having been widely adopted by many online services, including E-commerce, online news and social media sites. The key to a personalized recommender system is in modelling users' preference on items based on their past interactions (*e.g.*, ratings and clicks), known as collaborative filtering [31, 46]. Among the various collaborative filtering techniques, matrix factorization (MF) [14, 21] is the most popular one, which projects users and items into a shared latent space, using a vector of latent features to represent a user or an item. Thereafter a user's interaction on an item is modelled as the inner product of their latent vectors.

Popularized by the Netflix Prize, MF has become the *de facto* approach to latent factor model-based recommendation. Much research effort has been devoted to enhancing MF, such as integrating it with neighbor-based models [21], combining it with topic models of item content [38], and extending it to factorization machines [26] for a generic modelling of features. Despite the effectiveness of MF for collaborative filtering, it is well-known that its performance can be hindered by the simple choice of the interaction function — inner product. For example, for the task of rating prediction on explicit feedback, it is well known that the performance of the MF model can be improved by incorporating user and item bias terms into the interaction function¹. While it seems to be just a trivial tweak for the inner product operator [14], it points to the positive effect of designing a

Recommender System 이해


- 하이브리드(Hybrid) 기술 기술
 - 두 가지 이상의 알고리즘 혼합을 통한 하이브리드 형태









Recommender System 이해

- 기업에서의 추천시스템

- 당근마켓 - 같은 지역의 사람들이 어떤 상품을 관심있게 보는지 알려줌



이 상품과 함께 봤어요

 <p>리브맘 빈백 개당 500... 경기도 부천시 중1동 5,000원 관심 2 · 재팅 1</p>	 <p>후라이팬정리대 경기도 부천시 송내동 3,000원 관심 1 · 재팅 1</p>	 <p>이케아프랭클린바의... 경기도 부천시 중2동 20,000원 관심 4 · 재팅 1</p>
 <p>원조립 선반 "걸러는 ... 경기도 부천시 중1동 7,000원 관심 1 · 재팅 7</p>	 <p>그릇 경기도 부천시 원종2동 8,000원 관심 1 · 재팅 0</p>	 <p>선풍기 경기도 부천시 중1동 8,000원 관심 0 · 재팅 0</p>

Recommender System 이해

- 기업에서의 추천시스템
 - 카카오 브런치 - 고객이 읽은 글과 비슷한 내용의 글을 추천

실전 이탈 예측 모델링을 위한 세 가지 고려 사항 #1



Recommender System 이해

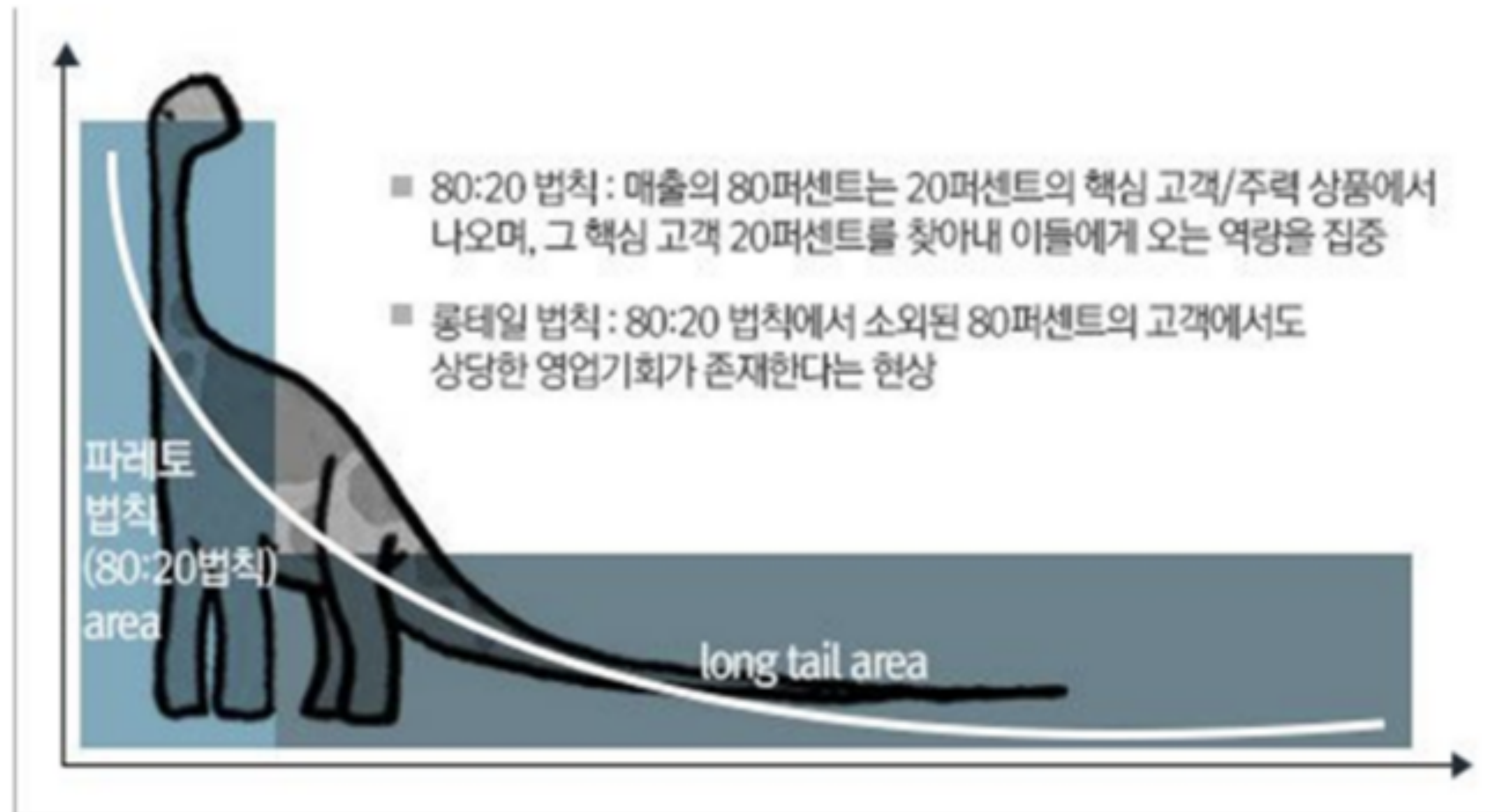
- 기업에서의 추천시스템
 - 파레토 법칙 VS 롱테일 법칙

✓ 파레토의 법칙

상위 20%가 80%의 가치를 창출한다.

✓ 롱테일의 법칙

하위 80%가 상위 20%의 가치보다 크다.



* 아마존 : 오프라인에서 팔지 않는 책은 온라인에서

추천하는 판매 전략

* 넷플릭스 : 개개인 맞춤 영화 추천으로 하위 80% 매출을 유도

Recommender System 이해

- 기업에서의 추천시스템
 - Netflix 영화 추천



- * 'CineMatch' 시스템
- * 넷플릭스에서 소비하는 영화 중 80%가 추천에 의한 것

Recommender System 이해

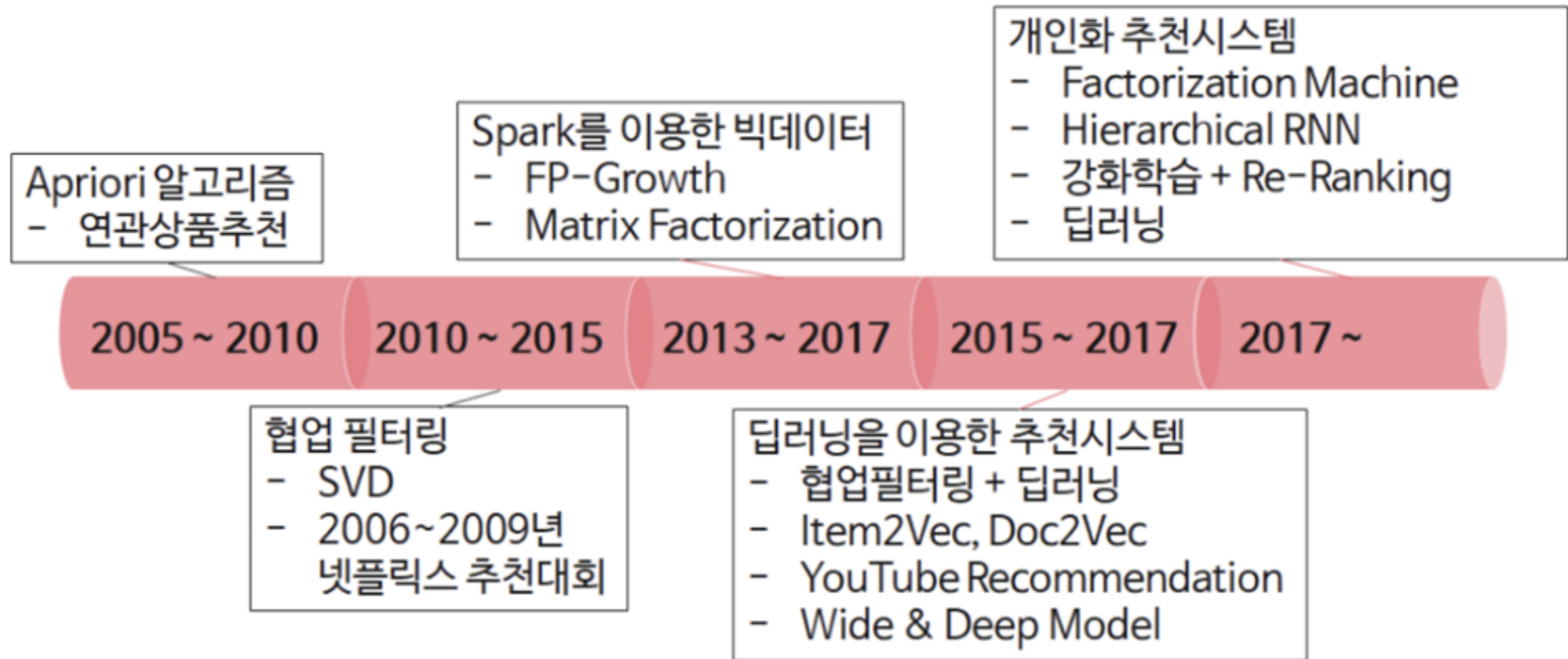
- 기업에서의 추천시스템
 - 아마존 상품 추천



- * 협업 필터링을 사용한 시스템 중 가장 오래된 시스템
- * 가장 성공적인 시스템 중 하나
- * 협업 필터링을 기반으로 하는 다양한 추천 기술을 결합한 하이브리드 기술 적용
- * 별점, 리뷰, 방문 페이지 정보, 장바구니 담은 정보 및 구매 정보 등을 활용
- * 대량의 데이터, 시스템이 처리하는 선호 정보의 종류, 제품 카테고리의 다양성 등 고려

Recommender System 이해

■ 추천시스템 Trend



감/사/합/니/다