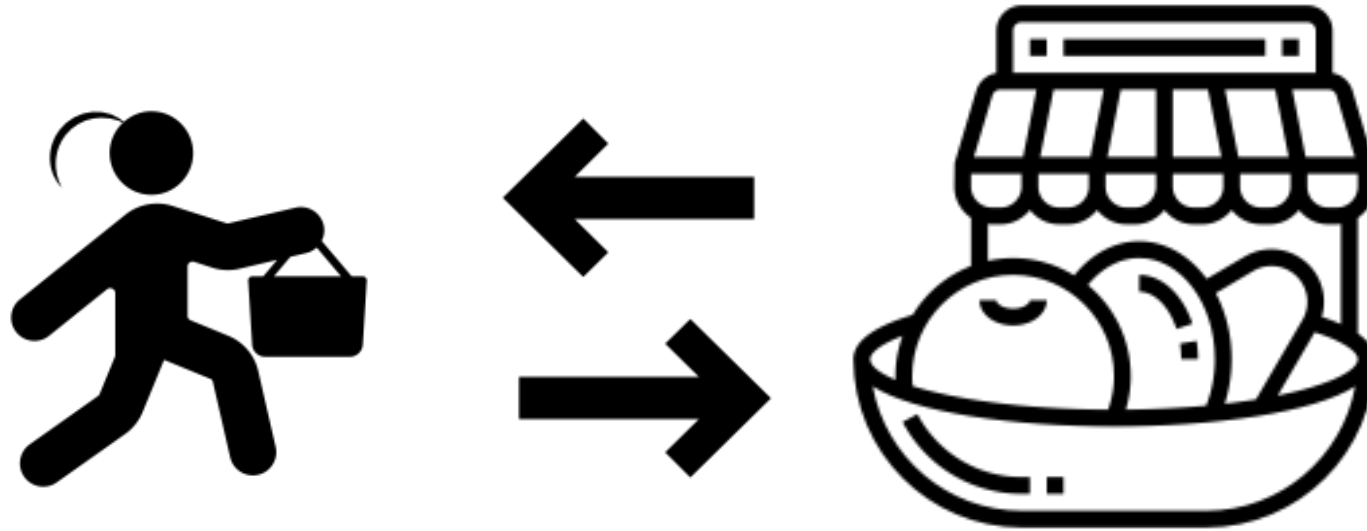


## 이 수업의 주제 : 추천

---



고객

상점

핵심 이슈 : 고객에게 **어떠한 제품을 권할 것인가?**

# 이 수업의 구성

---

## 1. 고객이 첫 방문을 했을 때

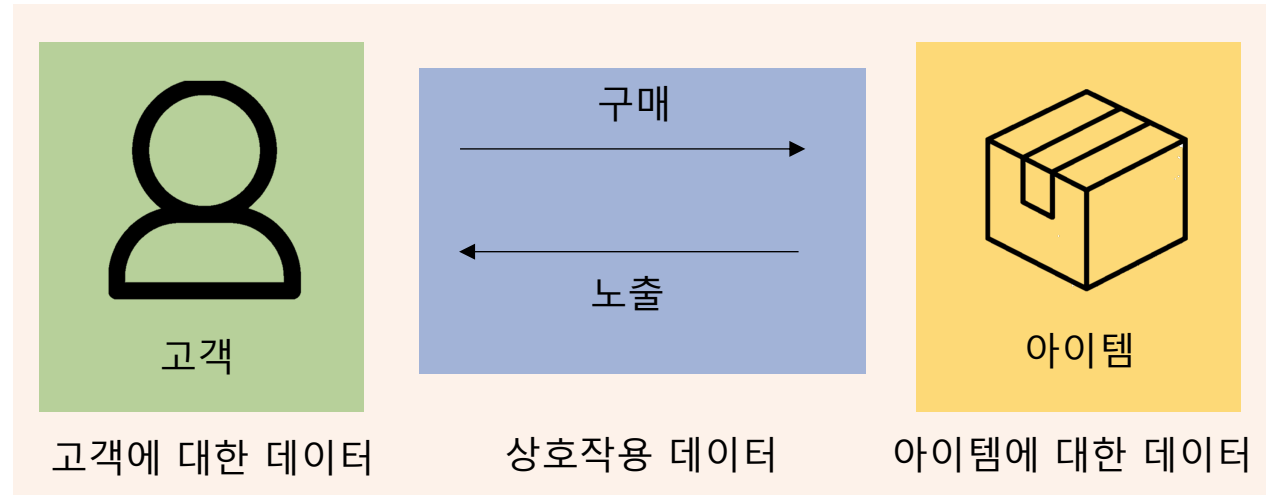
: 고객에 대한 정보가 없을 때

## 2. 고객이 재차 방문을 했을 때

: 고객의 이전 구매 정보가 있을 때

## 3. 고객이 단골이 되었을 때

: 고객 신원 정보가 있을 때



# 이 수업의 주요 토픽

---

## 1. 고객이 첫 방문을 했을 때

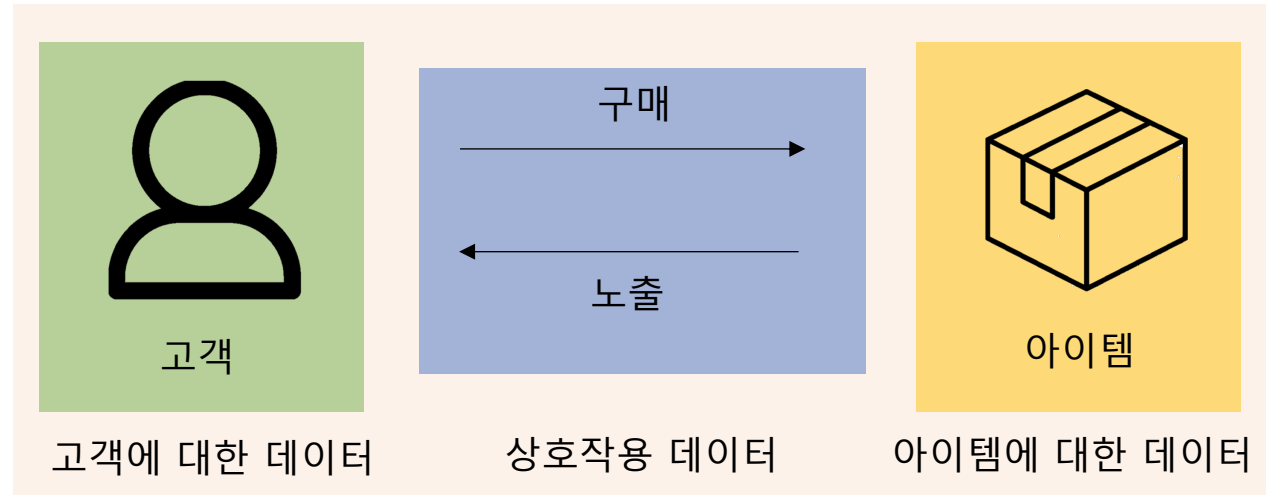
- 비개인화 추천
- 연관 분석

## 2. 고객이 재차 방문을 했을 때

- Collaborative Filtering

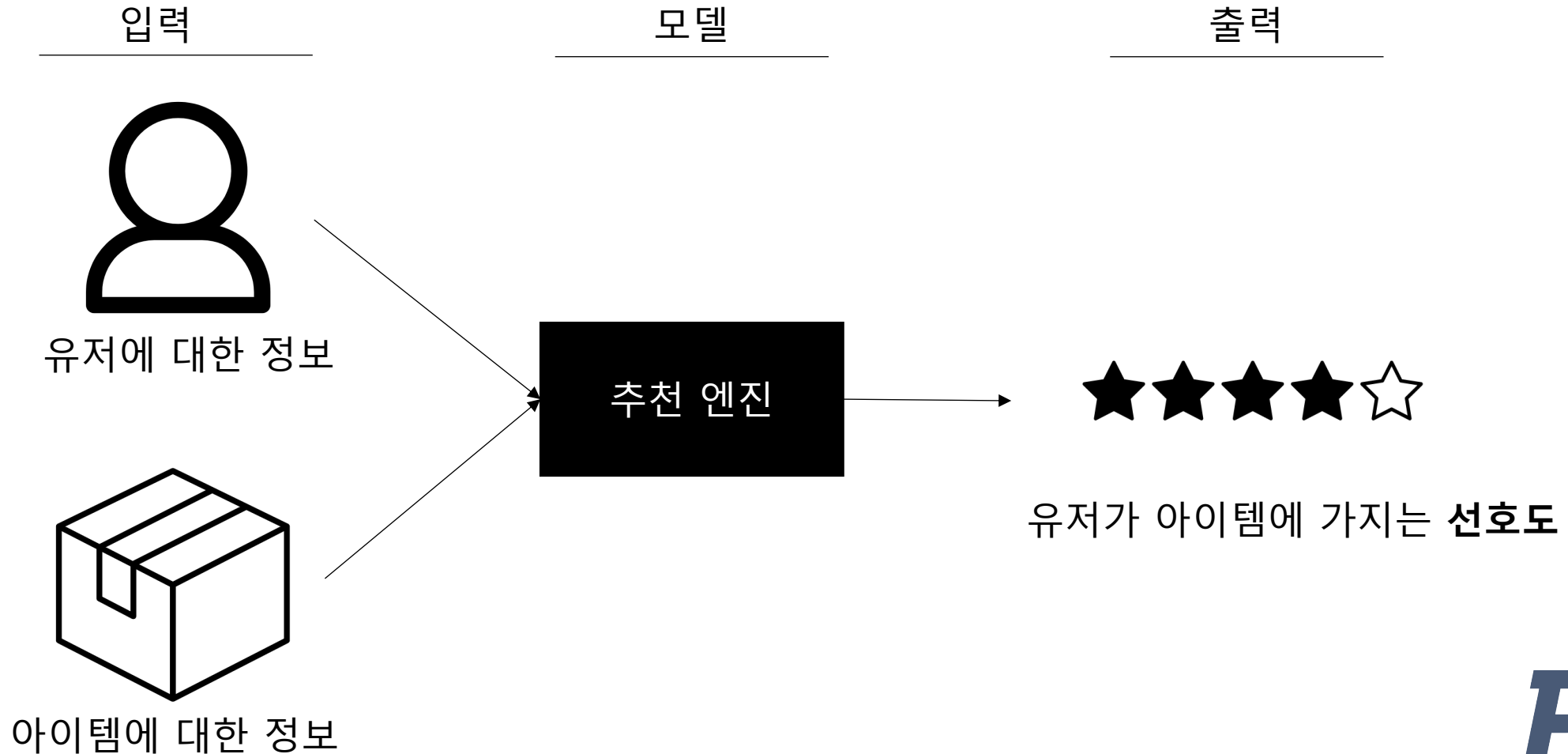
## 3. 고객이 단골이 되었을 때

- Factorization Machine



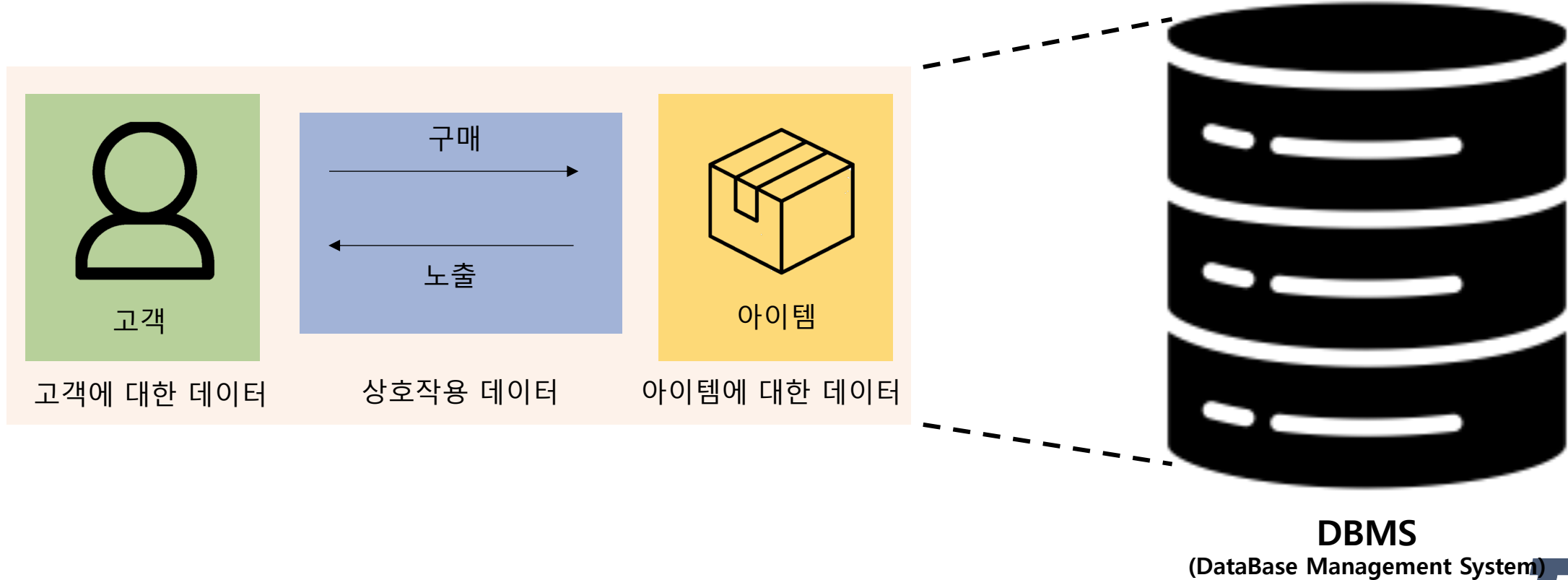
# 추천 엔진이란

추천 엔진의 과제 : 사용자가 얼마나 해당 아이템을 선호할지를 예측할 수 있는 프로그램을 구성하는 것\*



# 데이터베이스 <Database>

# 추천 시스템에서의 데이터



# 오늘 배우게 되는 SQL 구문들

---

원하는 데이터 셋을 가져오기 위한 방법들

**SELECT c1, c2 FROM t;**

Query data in columns c1, c2 from a table

**SELECT \* FROM t;**

Query all rows and columns from a table

**SELECT c1, c2 FROM t**

**WHERE condition;**

Query data and filter rows with a condition

**SELECT DISTINCT c1 FROM t**

**WHERE condition;**

Query distinct rows from a table

**SELECT c1, c2 FROM t**

**ORDER BY c1 ASC [DESC];**

Sort the result set in ascending or descending order

**SELECT c1, c2 FROM t**

**ORDER BY c1**

**LIMIT n OFFSET offset;**

Skip *offset* of rows and return the next n rows

**SELECT c1, aggregate(c2)**

**FROM t**

**GROUP BY c1;**

Group rows using an aggregate function

# 연관 분석

## <Association Analysis>



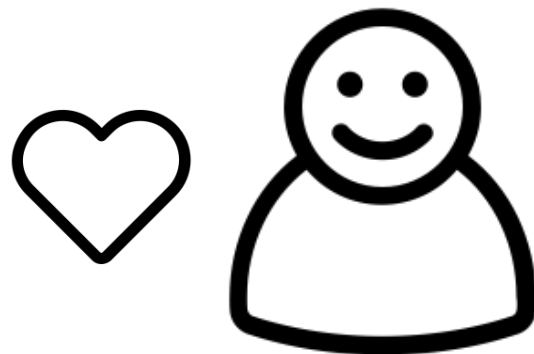
만약  $X$ 를 선호(구매)했다면,  
 $Y$ 도 선호(구매)할 것이다

의 연관 관계를 도출하는 것

스타워즈 2를 너무나 재밌게 본 유저에게  
어떤 영화를 추천하는 것이 좋을까?

(1) 지지도 :  $Support(X) = \frac{freq(X)}{N}$

: 전체 고객 중에서 영화 X를 선호하는 사람의 비율은?



지지도를 기준으로 추천할 경우,  
어떤 영화를 선호하던 항상 같은 영화(타이타닉)을  
추천하는 문제가 발생

지지도

4.1%

3.8%

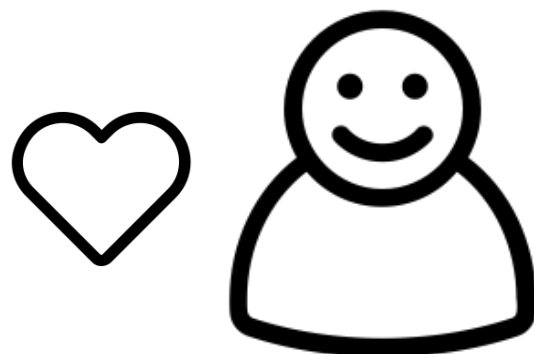
3.8%

10.0%

스타워즈 2를 너무나 재밌게 본 유저에게  
어떤 영화를 추천하는 것이 좋을까?

(2) 신뢰도 :  $confidence(X \rightarrow Y) = \frac{freq(X,Y)}{freq(X)}$

: 영화 X를 선호한 사람 중에서  
영화 Y를 선호하는 사람의 비율은?(조건부 확률)

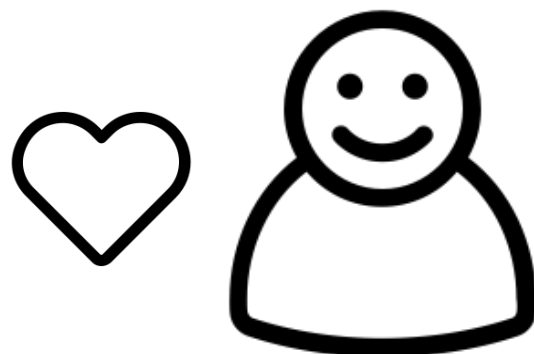


	지지도	신뢰도
스타워즈 3	4.1%	50.2%
스타 트렉	3.8%	17.0%
다들 좋아해요 그 사람들 좋아	3.8%	13.1%
리브 애츨리	10.0%	29.0%

신뢰도를 기준으로 추천할 경우,  
대다수 사람들이 선호하는 영화가  
소수의 사람들이 선호하는 영화보다 우선 추천됨

스타워즈 2를 너무나 재밌게 본 유저에게  
어떤 영화를 추천하는 것이 좋을까?

(3) 리프트 =  $\text{lift}(X \rightarrow Y) = \frac{\text{confidence}(X \rightarrow Y)}{\text{support}(Y)}$   
: 지지도 대비 신뢰도가 얼마나 올라갔는가?



리프트를 기준으로 추천할 경우,  
소수의 사람들이 선호하는 영화이더라도  
강한 연관관계가 있으면 영화가 우선 추천됨



	지지도	신뢰도	리프트
스타워즈 3: 반격의 함성	4.1%	50.2%	12.2
스타 트렉	3.8%	17.0%	4.4
다들 좋아해요	3.8%	13.1%	3.5
리브 애츨얼리	10.0%	29.0%	2.9

(1) 지지도 :  $Support(X) = \frac{freq(X)}{N}$

(2) 신뢰도 :  $confidence(X \rightarrow Y) = \frac{freq(X,Y)}{freq(X)}$

(3) 리프트 :  $lift(X \rightarrow Y) = \frac{confidence(X \rightarrow Y)}{support(Y)} = \frac{N * freq(X,Y)}{freq(X)freq(Y)}$

연관분석을 위해 알아야 하는 값 :

$freq(X), freq(Y), freq(X, Y)$

- $freq(X)$  : 아이템(X)을 선호(구매)한 경우의 수
- $freq(Y)$  : 아이템(Y)을 선호(구매)한 경우의 수
- $freq(X, Y)$  : 아이템(X,Y)을 동시에 선호(구매)한 경우의 수

## 연관분석의 단위 : 장바구니

연관분석은 **장바구니**,  
즉 고객이 동시에 구매하거나 선호하는 제품군에 대한 정보가 필요

Case ) 제품이 a, b, c, d, e 가 있는 경우

	장바구니 (고객이 동시에 구매한 제품 목록)
0	{b, c}
1	{b, d, a}
2	{c, e}
3	{c, d, e}
4	{c, d}
5	{b, d}
6	{b, c, e}
7	{d, e}
8	{b, c, d}
9	{c, e}
10	{b, e}
11	{b, c, d, e}

## 연관분석의 어려움, 너무 많은 연관 관계

연관분석을 진행하기 위해서는 각 아이템의 동시에 등장한 횟수를 계속 세어야 함  
이는 엄청나게 많은 경우의 수가 존재

	장바구니 (고객이 함께 선호 혹은 구매한 제품군)
0	{b, c}
1	{b, d, a}
2	{c, e}
3	{c, d, e}
4	{c, d}
5	{b, d}
6	{b, c, e}
7	{d, e}
8	{b, c, d}
9	{c, e}
10	{b, e}
11	{b, c, d, e}



$\{x\} \rightarrow \{y\}$  : 만약  $x$ 를 구매하였으면,  $y$ 를 구매할 것이다

$\{a\} \rightarrow \{b\}$	$\{c\} \rightarrow \{a\}$	$\{e\} \rightarrow \{a\}$	$\{a,b,c\} \rightarrow \{d\}$	$\{b, c\} \rightarrow \{d\}$
$\{a\} \rightarrow \{c\}$	$\{c\} \rightarrow \{b\}$	$\{e\} \rightarrow \{b\}$	$\{a,b,c\} \rightarrow \{e\}$	
$\{a\} \rightarrow \{d\}$	$\{c\} \rightarrow \{d\}$	$\{e\} \rightarrow \{c\}$	$\{a,b,d\} \rightarrow \{c\}$	
$\{a\} \rightarrow \{e\}$	$\{c\} \rightarrow \{e\}$	$\{e\} \rightarrow \{d\}$	$\{a,b,d\} \rightarrow \{e\}$	
$\{b\} \rightarrow \{a\}$	$\{d\} \rightarrow \{a\}$	$\{a,b\} \rightarrow \{c\}$	$\{a,c\} \rightarrow \{b\}$	
$\{b\} \rightarrow \{c\}$	$\{d\} \rightarrow \{b\}$	$\{a,b\} \rightarrow \{d\}$	$\{a,c\} \rightarrow \{d\}$	
$\{b\} \rightarrow \{d\}$	$\{d\} \rightarrow \{c\}$	$\{a,b\} \rightarrow \{e\}$	$\{a,c\} \rightarrow \{e\}$	
$\{b\} \rightarrow \{e\}$	$\{d\} \rightarrow \{e\}$	$\{b,c\} \rightarrow \{a\}$	$\{b,c\} \rightarrow \{e\}$	

: 너무 많은 연관 관계가 존재

## 빈발집합 찾기 알고리즘 : 아프리오리

모든 연관 관계를 파악할 수 없기 때문에, "**자주 등장한**" 연관 관계만을 파악하자 -> 아프리오리의 아이디어

	장바구니 (고객이 함께 선호 혹은 구매한 제품군)
0	{b, c}
1	{b, d, a}
2	{c, e}
3	{c, d, e}
4	{c, d}
5	{b, d}
6	{b, c, e}
7	{d, e}
8	{b, c, d}
9	{c, e}
10	{b, e}
11	{b, c, d, e}



## 빈발집합 찾기 알고리즘 : 아프리오리

1. 제품별로 각각 장바구니에 몇번씩 등장했는지를 셈

장바구니 (고객이 함께 선호 혹은 구매한 제품군)

0	{b, c}	a	b	c	d	e					
1	{b, d, a}										
2	{c, e}										
3	{c, d, e}	ab	a,c	a,d	a,e	b,c	b,d	b,e	c,d	c,e	d,e
4	{c, d}										
5	{b, d}										
6	{b, c, e}	a, b,c	a,b, d	a,b, e	a,c, d	a,c, e	a,d, e	b,c, d	b,c, e	b, d,e	c,d, e
7	{d, e}										
8	{b, c, d}										
9	{c, e}										
10	{b, e}										
11	{b, c, d, e}										

1

a

7

b

8

c

7

d

7

e

ab

a,c

a,d

a,e

b,c

b,d

b,e

c,d

c,e

d,e

a, b,c

a,b, d

a,b, e

a,c, d

a,c, e

a,d, e

b,c, d

b,c, e

b, d,e

c,d, e

ab cd

ab ce

ab de

ac de

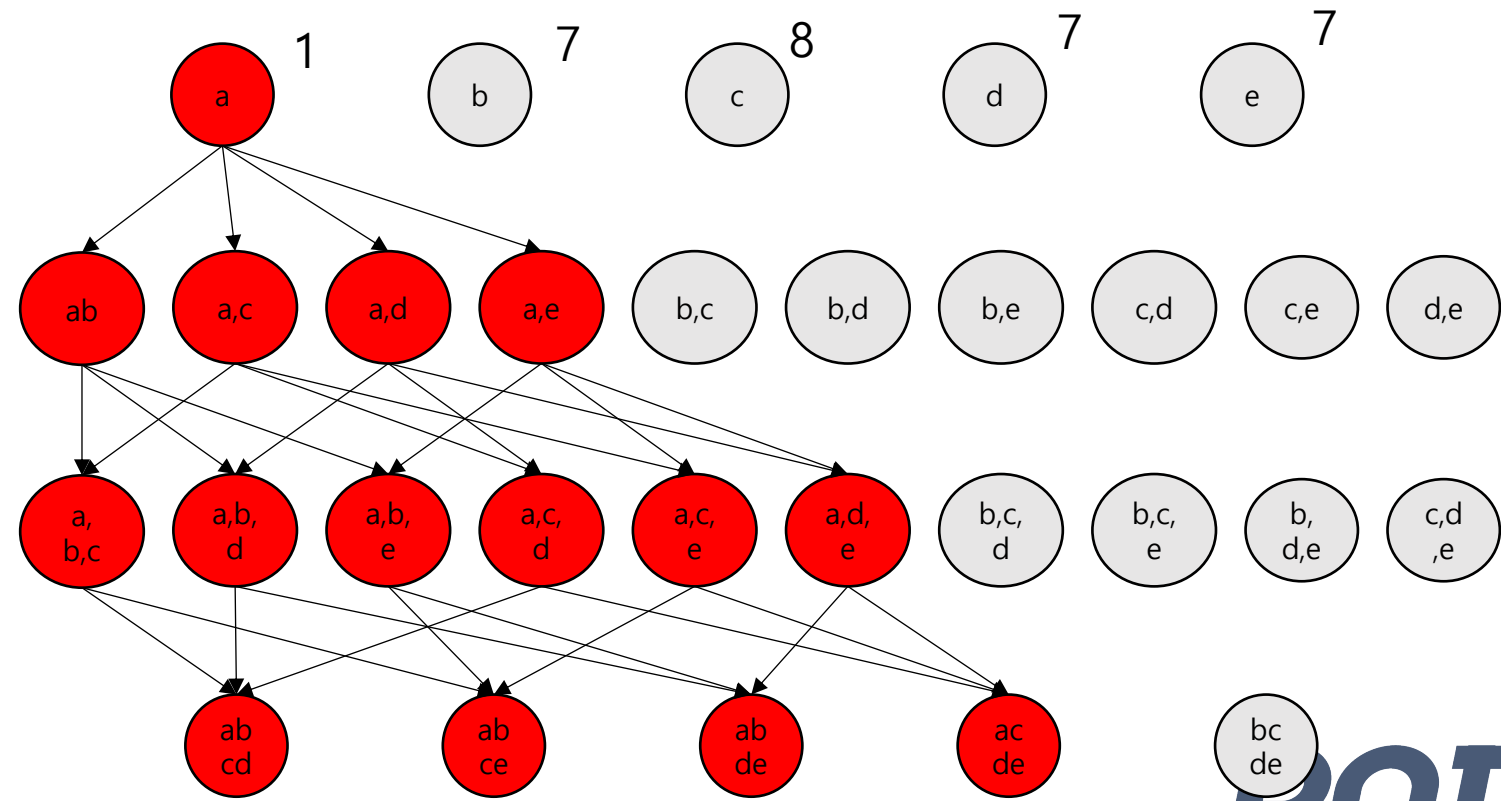
bc de

PAI

## 빈발집합 찾기 알고리즘 : 아프리오리

2. 너무 적게 등장한 연관관계의 경우, 해당 아이템 집합을 포함한 모든 집합은 카운트하지 않음

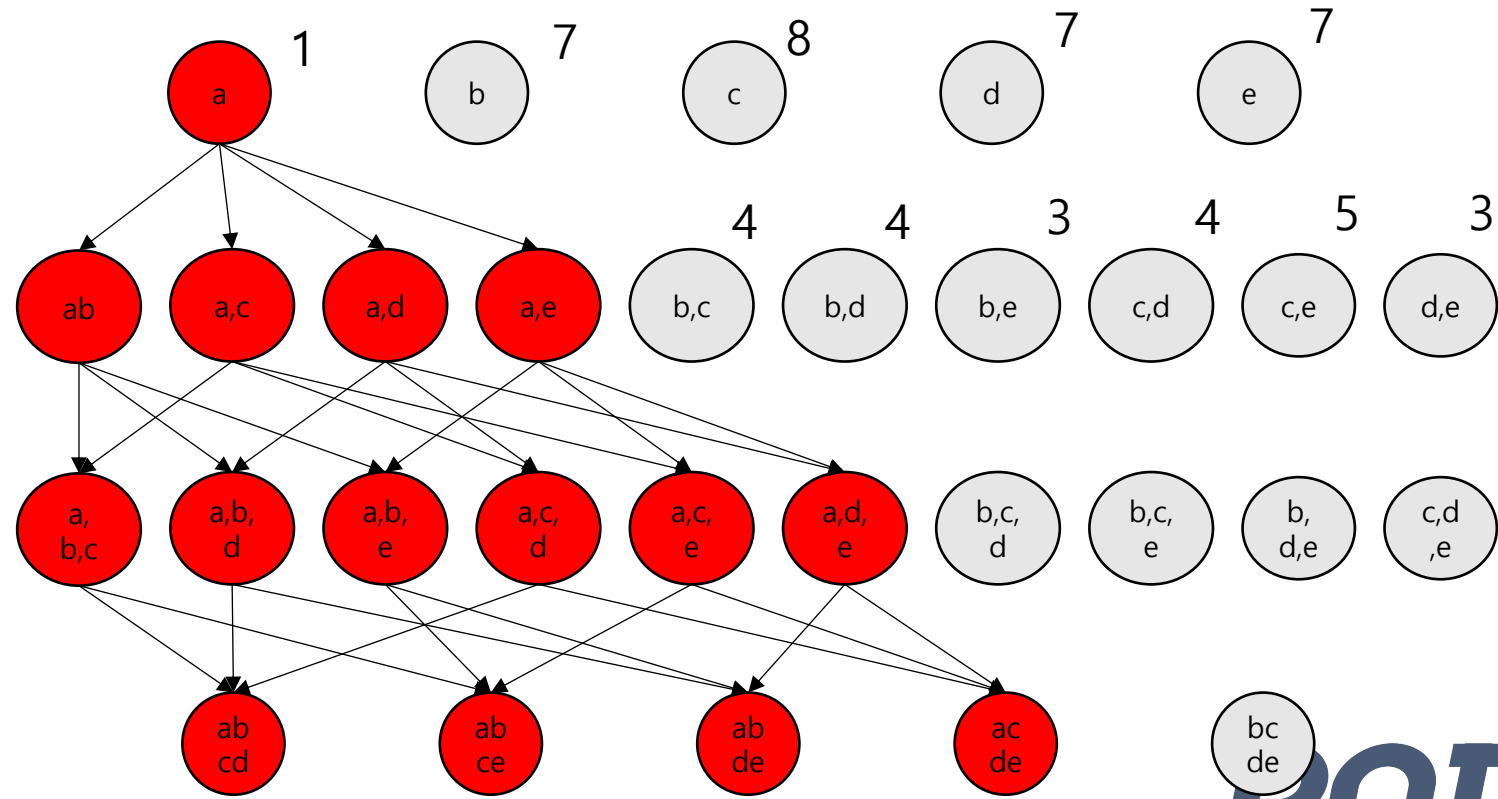
	장바구니 (고객이 함께 선호 혹은 구매한 제품군)
0	{b, c}
1	{b, d, a}
2	{c, e}
3	{c, d, e}
4	{c, d}
5	{b, d}
6	{b, c, e}
7	{d, e}
8	{b, c, d}
9	{c, e}
10	{b, e}
11	{b, c, d, e}



## 빈발집합 찾기 알고리즘 : 아프리오리

### 3. 생략되고 남은 것들에 한에 카운트를 함

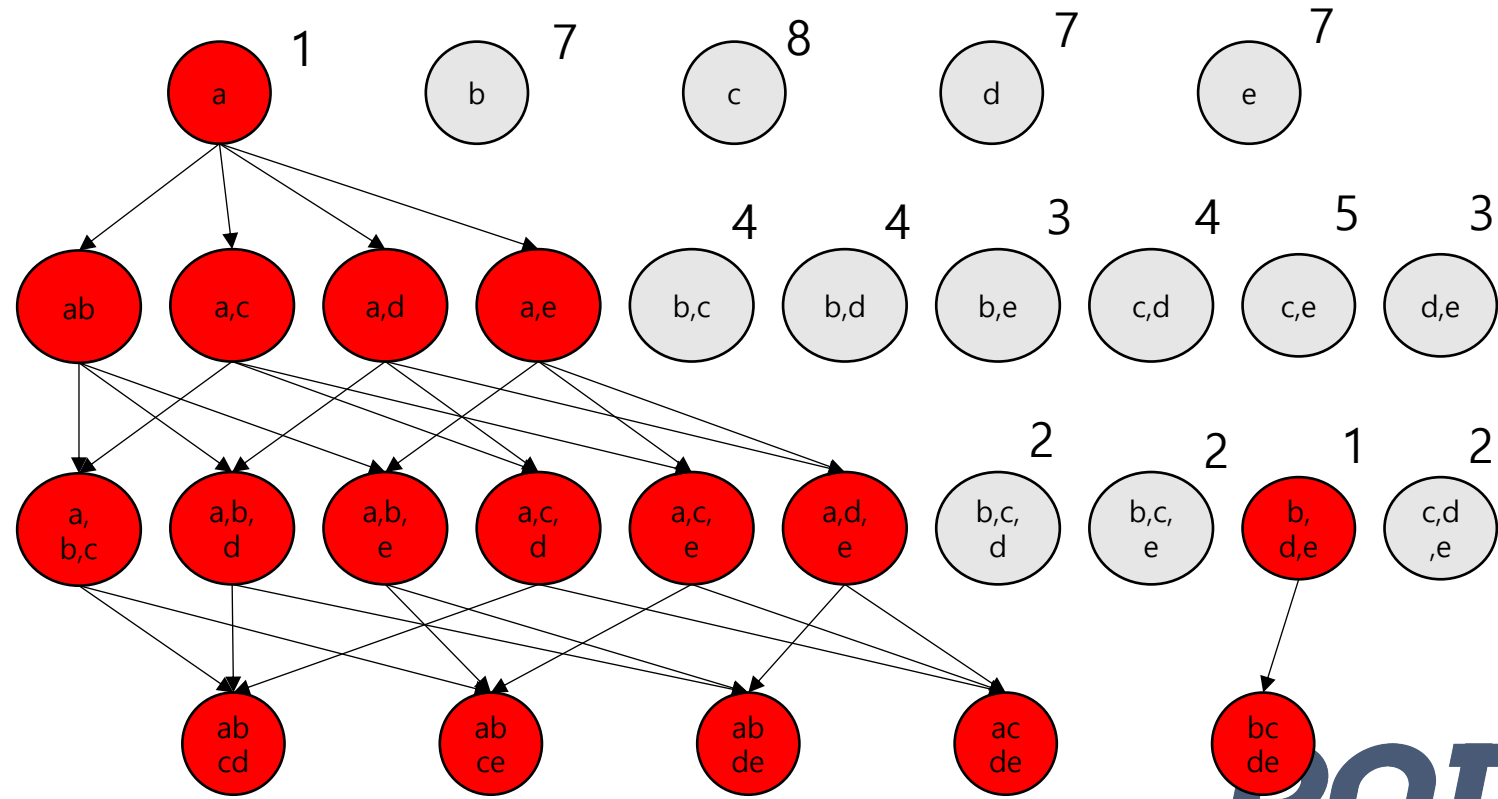
	장바구니 (고객이 함께 선호 혹은 구매한 제품군)
0	{b, c}
1	{b, d, a}
2	{c, e}
3	{c, d, e}
4	{c, d}
5	{b, d}
6	{b, c, e}
7	{d, e}
8	{b, c, d}
9	{c, e}
10	{b, e}
11	{b, c, d, e}



## 빈발집합 찾기 알고리즘 : 아프리오리

### 3. 생략되고 남은 것들에 한에 카운트를 함

	장바구니 (고객이 함께 선호 혹은 구매한 제품군)
0	{b, c}
1	{b, d, a}
2	{c, e}
3	{c, d, e}
4	{c, d}
5	{b, d}
6	{b, c, e}
7	{d, e}
8	{b, c, d}
9	{c, e}
10	{b, e}
11	{b, c, d, e}



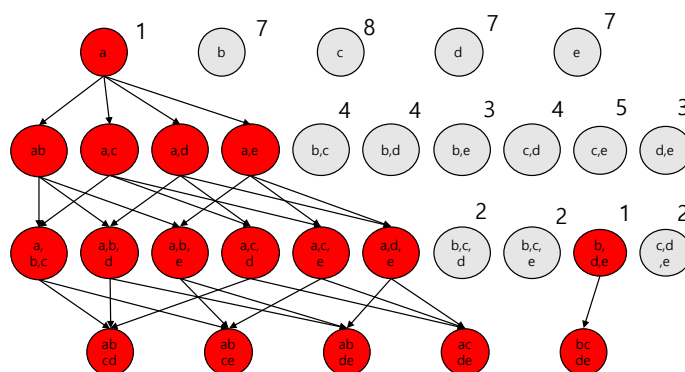
## 빈발집합 찾기 알고리즘 : 아프리오리

탐색할 아이템 집합이 없어지면, 종료 후 빈발집합 도출

빈발집합 도출

	장바구니 (고객이 함께 선호 혹은 구매한 제품군)
0	{b, c}
1	{b, d, a}
2	{c, e}
3	{c, d, e}
4	{c, d}
5	{b, d}
6	{b, c, e}
7	{d, e}
8	{b, c, d}
9	{c, e}
10	{b, e}
11	{b, c, d, e}

아프리오리 알고리즘



빈발집합	Freq
{b}	7
{c}	8
{d}	7
{e}	7
{b,c}	4
{b,d}	4
{b,e}	3
{c,d}	4
{c,e}	5
{d,e}	3
{b,c,d}	2
{b,c,e}	2
{c,d,e}	2

## 빈발집합을 통한 연관 분석

---

빈발 집합이 도출되고 나면 그 후 바로 연관 관계를 수식으로 계산 가능

빈발집합	Freq
{b}	7
{c}	8
{d}	7
{e}	7
{b,c}	4
{b,d}	4
{b,e}	3
{c,d}	4
{c,e}	5
{d,e}	3
{b,c,d}	2
{b,c,e}	2
{c,d,e}	2

연관 관계 (  $b \rightarrow c$  )

: 만약 b를 구매하였을 때, c도 동시에 구매할까요?

(1) 지지도 :  $Support(c) = \frac{8}{12}$

(2) 신뢰도 :  $confidence(b \rightarrow c) = \frac{freq(b,c)}{freq(b)} = \frac{4}{7}$

(3) 리프트 :  $lift(b \rightarrow c) = \frac{confidence(b \rightarrow c)}{support(c)} = \frac{14}{12}$

## 빈발집합을 통한 연관 분석

---

빈발 집합이 도출되고 나면 그 후 바로 연관 관계를 수식으로 계산 가능

빈발집합	Freq
{b}	7
{c}	8
{d}	7
{e}	7
{b,c}	4
{b,d}	4
{b,e}	3
{c,d}	4
{c,e}	5
{d,e}	3
{b,c,d}	2
{b,c,e}	2
{c,d,e}	2

연관 관계 (  $c \rightarrow e$  )

: 만약  $c$ 를 구매하였을 때,  $e$ 도 동시에 구매할까요?

(1) 지지도 :  $Support(e) = \frac{7}{12}$

(2) 신뢰도 :  $confidence(c \rightarrow e) = \frac{freq(c,e)}{freq(c)} = \frac{5}{8}$

(3) 리프트 :  $lift(c \rightarrow e) = \frac{confidence(c \rightarrow e)}{support(c)} = \frac{60}{56}$

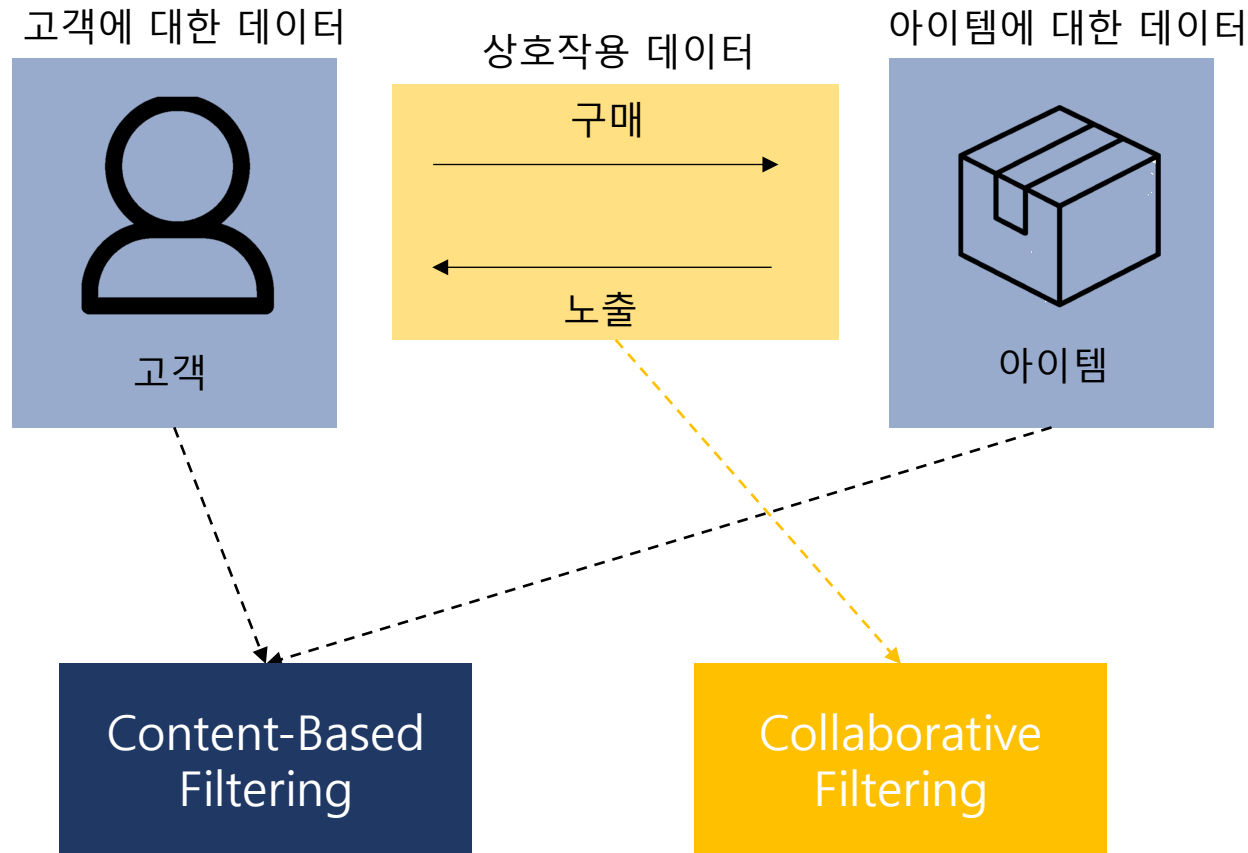
# 협업 필터링

## <Collaborative Filtering>



# 추천 시스템의 알고리즘

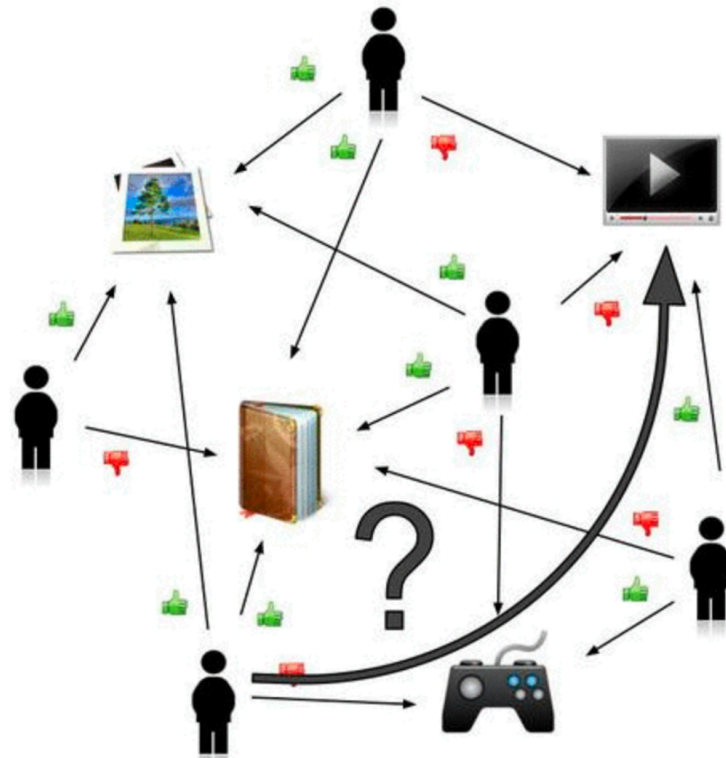
추천 시스템은 크게 콘텐츠 기반 추천과 협업 필터링 기반 추천으로 나뉘어짐



# 협업 필터링이란?

## 핵심 아이디어

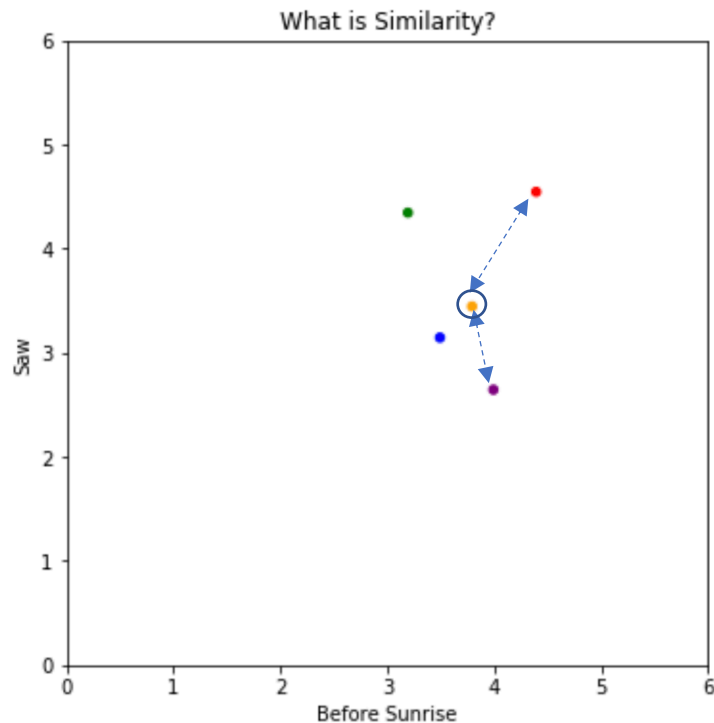
만약 두 명의 사용자가 **유사한** 관심사를 가지고 있다면,  
그들은 미래에도 **유사한** 취향을 가질 것이다.



# 협업 필터링 연산의 핵심 : 유사도

## 핵심 아이디어

만약 두 명의 사용자가 **유사한** 관심사를 가지고 있다면,  
그들은 미래에도 **유사한** 취향을 가질 것이다.



## 유사도 계산

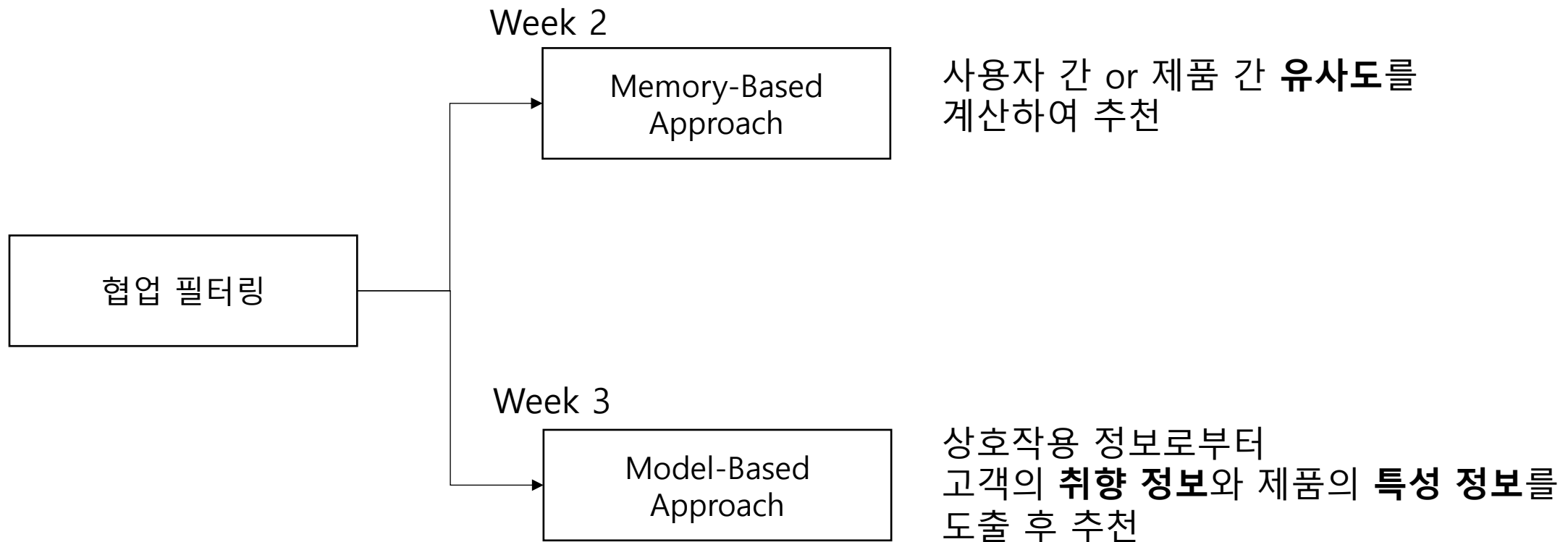
노란색 유저는

\* 빨간색 유저와 유사할까?

\* 보라색 유저와 유사할까?

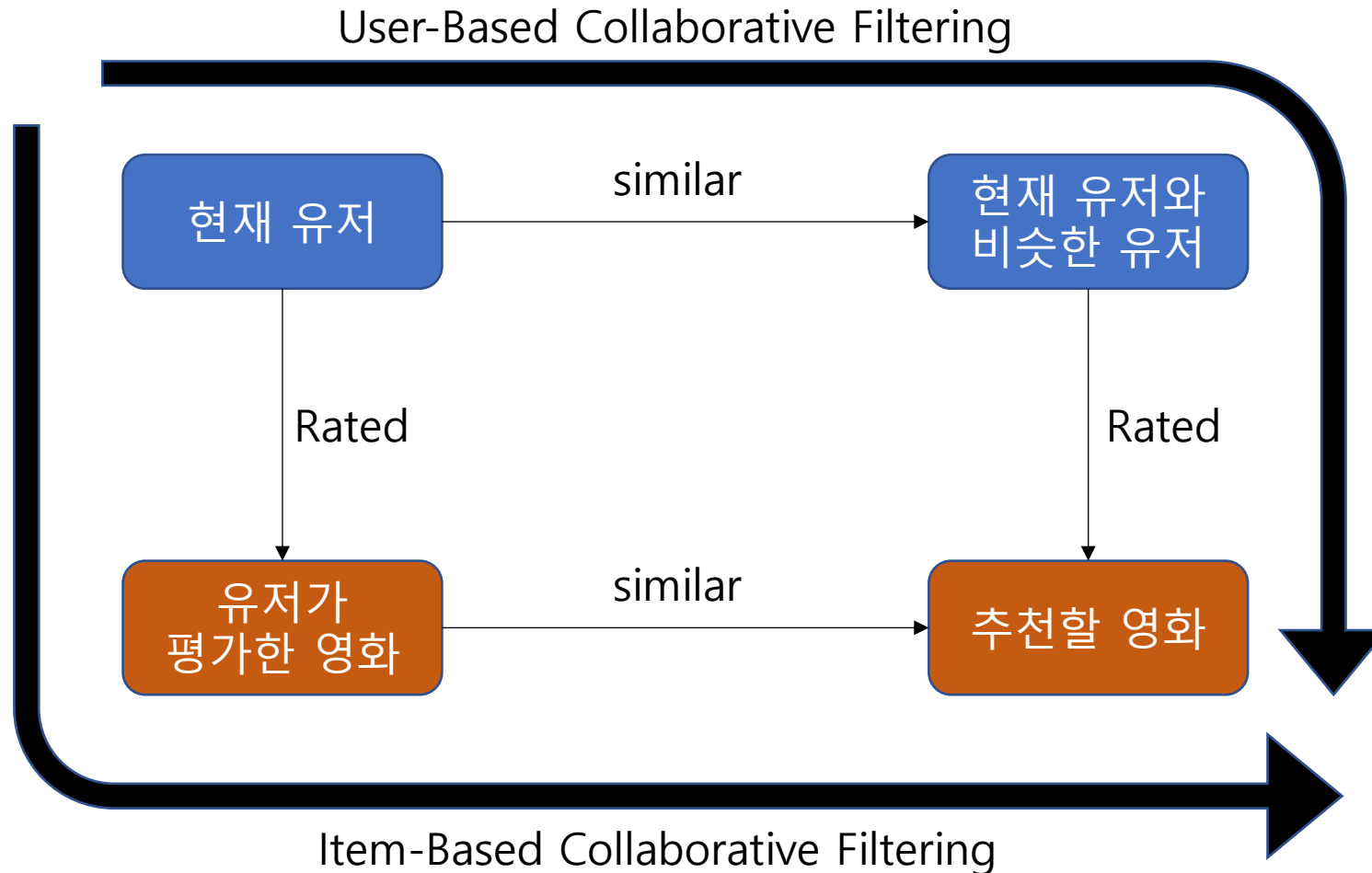
# 협업 필터링의 종류

협업 필터링에는 Memory-Based Approach와 Model-Based Approach로 나뉘어짐



# Memory-Based 협업 필터링의 종류

Memory-Based 협업 필터링은 크게  
User-Based Collaborative Filtering과 Item-Based Collaborative Filtering으로 나뉘어짐






















User-Item Matrix

영화 유저	영화 이름	유저 1	유저 2	유저 3	유저 4
	영화 1	👍	👎	👎	
	영화 2	👍	👍	👍	
	영화 3	👍	👎	👍	
	영화 4	👍	👎	👍	









Item Similarity Matrix

영화 유저	영화 1	영화 2	영화 3
영화 1	1	0.25	0.75
영화 2	0.25	1	0.5
영화 3	0.75	0.5	1

User-Item Matrix

영화 유저			
			
			
			
			

User Similarity Matrix

영화 유저				
	1	0.33	0.66	0.66
	0.33	1	0.66	0.66
	0.66	0.66	1	1
	0.66	0.66	1	1

### Item-User Matrix

영화 \ 유저				
영화				

●  
행렬 곱

### User-Item Matrix

영화 \ 유저			
유저			

=

### Similarity Matrix

	1	0.25	0.75
	0.25	1	0.5
	0.75	0.5	1