

Chap10-2.

추천시스템 모형

작성자 : 김진성



목차

- 1) 추천시스템 개요
- 2) 추천시스템 알고리즘
- 3) 추천시스템 사례
- 4) 유사도 계산
- 5) 특정 사용자 유사도 평점으로 추천 받기
- 6) 특이값 분해(SVD)



1) 추천 시스템 개요

- 정보 필터링 (IF) 기술의 일종으로, 특정 사용자가 관심을 가질만한 정보 (영화, 음악, 책, 뉴스, 이미지, 웹 페이지 등)를 추천하는 시스템

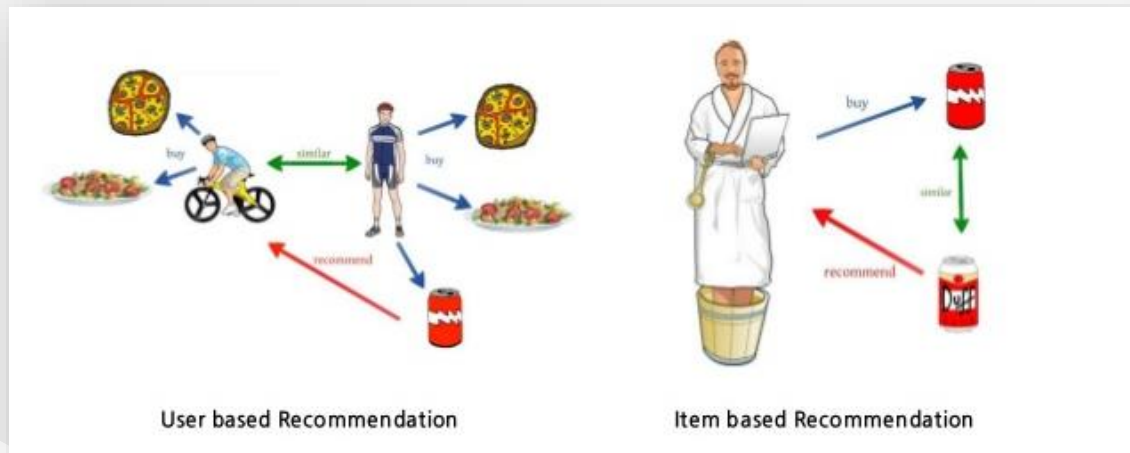
- 추천 알고리즘
 1. 협업 필터링(Collaborative Filtering : CF)
 2. 내용기반 필터링(Content-Based Filtering : CB)
 3. 지식기반 필터링(Knowledge-Based Filtering : KB)



2) 추천 시스템 알고리즘

1. 협업 필터링(Collaborative Filtering : CF)

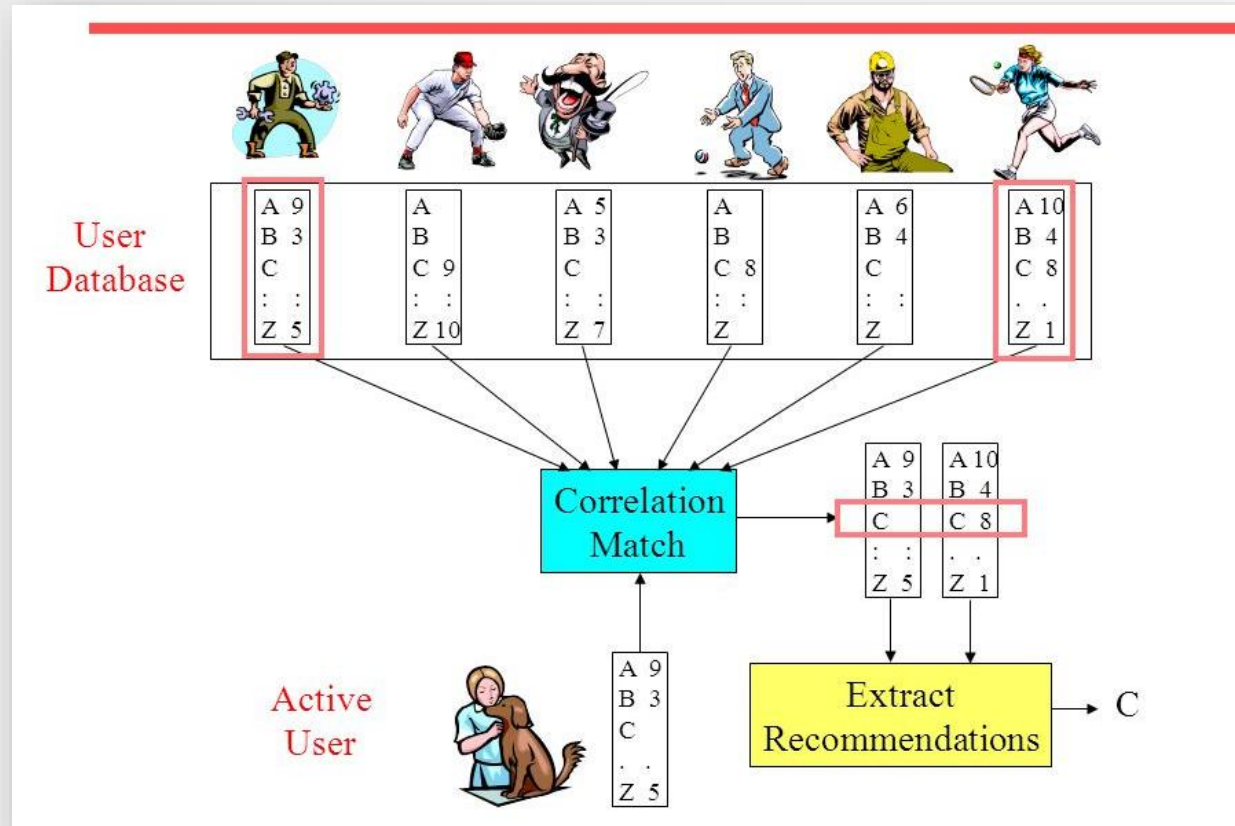
- ✓ 구매/소비 패턴이 비슷한 사용자를 한 집단으로 보고 그 집단에 속한 소비자들의 취향을 추천하는 방식
- ✓ UBCF(User Based CF)
 - 패턴이 비슷한 사용자를 기반으로 상품(Item) 추천 방식
- ✓ IBCF(Item Based CF)
 - 상품(Item)을 기반으로 연관성이 있는 상품(Item) 추천 방식





2) 추천 시스템 알고리즘

- 사용자 기반 협업 필터링(UBCF) 예





2) 추천 시스템 알고리즘

- 사용자 기반 협업 필터링(UBCF) 예

5명의 사용자가 5개의 영화에 대해서 평점을 기록한 matrix A
A 사용자와 가장 유사한 사용자를 선정하여 movie4와 movie5 중 하나의 영화 추천받기

user	movie1	movie2	movie3	movie4	movie5
a	2	5	3	NA	NA
b	4	4	3	5	1
c	1	5	4	NA	5
d	3	5	3	2	5
e	4	5	3	4	NA

	a	b	c	d	e
a	1.0000000	0.1889822	0.8910421	0.9449112	0.6546537
b	0.1889822	1.0000000	-0.2773501	0.5000000	0.8660254
c	0.8910421	-0.2773501	1.0000000	0.6933752	0.2401922
d	0.9449112	0.5000000	0.6933752	1.0000000	0.8660254
e	0.6546537	0.8660254	0.2401922	0.8660254	1.0000000

유사도
계산



2) 추천 시스템 알고리즘

- 사용자 기반 협업 필터링(BCF) 예

아래는 6명의 사용자가 8개의 아이템에 대해 평가한 매트릭스이다.
사용자 x와 가장 유사한 사용자를 알아내기 위해서 사용자 간의 거리를 측정해서 가장 가까운 3명의 사용자들을 찾아서 필터링한 결과

	아이템1	아이템2	아이템3	아이템4	아이템5	아이템6	아이템7	아이템8
사용자1	?	4.0	4.0	2.0	1.0	2.0	?	?
사용자2	3.0	?	?	?	5.0	1.0	?	?
사용자3	3.0	?	?	3.0	2.0	2.0	?	3.0
사용자4	4.0	?	?	2.0	1.0	1.0	2.0	4.0
사용자5	1.0	1.0	?	?	?	?	?	1.0
사용자6	?	1.0	?	?	1.0	1.0	?	1.0

사용자x	?	?	4.0	3.0	?	1.0	?	5.0
------	---	---	-----	-----	---	-----	---	-----

사용자x (예상 평점)	3.5	4.0			1.3		2.0	
-----------------	-----	-----	--	--	-----	--	-----	--

TopN=3



2) 추천 시스템 알고리즘

- 아이템 기반 협업 필터링(UBCF) 예

사용자(행), 아이템(열)의 matrix를 대상으로 아이템 A와 아이템 B의 벡터가 비슷하다면 A를 샀을 때 B를 추천해주는 알고리즘
예를 들어, 다음과 같은 3명의 사용자를 대상으로 각 상품을 구매(1)와 미구매(0)를 나타내는 matrix가 있다고 가정한다.

사용자	커피	녹차	카페라테
홍길동	1	0	1
이순신	1	1	1
유관순	0	0	1

커피(1,1,0) 벡터, 녹차(0,1,0) 벡터, 카페라테(1,1,1) 벡터를 가지므로
커피와 카페라테는 비슷한 벡터 구조를 갖는다. 따라서 커피를 샀을 때
카페라테를 사라고 추천할 수 있다. 녹차는 커피와 카페라테와 다르기
때문에 유사도가 낮아 추천해주지 않을 것이다.
이러한 과정을 보통 Item based Collaborative Filtering이라고 한다.



2) 추천 시스템 알고리즘

2. 내용기반 필터링(Content-Based Filtering : CB)

- ✓ 소비자가 소비하는 제품 중 텍스트 정보가 많은 제품 대상
- ✓ 뉴스, 책 등 텍스트의 내용을 분석해서 추천하는 방법
- ✓ 텍스트 중에서 형태소(명사, 동사 등)를 분석하여 핵심 키워드를 분석하는 기술이 핵심

3. 지식기반 필터링(Knowledge-Based Filtering : CB)

- ✓ 특정 분야에 대한 전문가의 도움을 받아서 그 분야에 대한 전체적인 지식구조를 만들고 이를 활용하는 방법



3) 추천 시스템 사례

- 넷플릭스(www.netflixx.com) 사례
 - ✓ 전 세계 3,000만명이 넘는 고객 대상 온라인 영화 대여회사
 - ✓ 고객의 영화 평가를 바탕으로 특정 고객에게 영화 추천 서비스
 - ✓ 고객 이탈률(churn rate) 4% 이하
- 아마존(amazon.com) 사례
 - ✓ 협업필터링 알고리즘 기반 추천 시스템 적용
 - ✓ 제품 웹 페이지 방문 기록, 쇼핑장바구니 기능, 구매 상품 선호 등 다양한 정보를 토대로 추천 시스템 구현



3) 추천 시스템 사례

- 아마존(amazon.com) 사례 – Item 기반 추천시스템 예

Amazon.com: Tablet Stand-Adjustable-Lamical-reader/dp/B01BD1V1KY/ref=br_msw_pdt_37_encoding=UTF&smid=A2072W1V15W1N&pf_rd_m=ATVPDKIKX0DER&pf_rd_s=&pf_rd_r=SC250Y0ZTBFFG5XYA4358&pf_rd_t=36701&pf_rd_p=...

Lamical Tablet Stand Adjustable, Lamical iPad Stand: Desktop Stand Holder Dock for new iPad 2017 Pro 9.7, 10.5, Air mini 2.3.4, Kindle, Nexus, Accessories, Tab, E-reader, other Tablets (4-13 inch) - Silver

Price: \$49.99
 With Deal: \$14.39
 You Save: \$35.60 (79%)

Deal is 93% Claimed
 This item ships to: Korea, Republic of (South Korea). Want it Wednesday, Jan. 3? Order within 11 hrs 6 mins and choose AmazonGlobal Priority Shipping at checkout. Learn more

Color: Silver

Customers who bought this item also bought

Sponsored products related to this item (What's this?)

What other items do customers buy after viewing this item?

재5장 Recommende... R | 재5장 Recommende... R | chap05_Recommende...txt | chap05_Recommende...txt | chap05_Recommende...txt | chap05_Recommende...txt | chap05_Recommende...txt | 전체 보기



4) 유사도 계산

- 협업 필터링의 유사도(Similarity) 계산 방법
 - ✓ 상관계수(Correlation coefficient) 유사도 : 피어슨 상관계수 이용
 - ✓ 코사인(Cosine) 유사도 : 두 벡터 사이의 각도
 - ✓ Jaccard 유사도 : 이진화 자료(binary data) 대상 유사도 계산
 - ✓ 유클리드 거리계산법 : 거리 기반 유사도 계산





5) 특정 사용자 유사도 평점으로 추천 받기

[시나리오]

Toby는 3편의 영화를 관람하고, 3편의 영화는 미 관람한 상태이다.
미 관람 영화 3편 중 자기 취향에 가장 적합한 영화를 추천 받기를 원한다.

[알고리즘]

- Toby 영화평점과 다른 사람 영화평점 간의 유사도를 계산하여 영화 추천받기
 1. Toby와의 유사도 계산(피어슨 상관계수)
 2. 유사도 평점 계산 = Toby 미관람 영화평점 * Toby와의 유사도
 3. Toby에게 영화 추천 = 유사도 평점 합계 / Toby와의 유사도 합계



movies

	critic	title	rating
1	Jack	Lady	3.0
2	Jack	Snakes	4.0
3	Jack	You Me	3.5
4	Jack	Superman	5.0
5	Jack	The Night	3.0
6	Mick	Lady	3.0
7	Mick	Snakes	4.0
8	Mick	Just My	2.0
9	Mick	Superman	3.0
10	Mick	You Me	2.0
11	Mick	The Night	3.0
12	Claudia	Snakes	3.5
13	Claudia	Just My	3.0
14	Claudia	You Me	2.5
15	Claudia	Superman	4.0
16	Claudia	The Night	4.5
17	Lisa	Lady	2.5
18	Lisa	Snakes	3.5
19	Lisa	Just My	3.0
20	Lisa	Superman	3.5
21	Lisa	The Night	3.0
22	Lisa	You Me	2.5
23	Toby	Snakes	4.5
24	Toby	Superman	4.0
25	Toby	You Me	1.0
26	Gene	Lady	3.0
27	Gene	Snakes	3.5
28	Gene	Just My	1.5
29	Gene	Superman	5.0
30	Gene	You Me	3.5
31	Gene	The Night	3.0

dcast(movies, title~critic, sum)

	title	Claudia	Gene	Jack	Lisa	Mick	Toby
1	Just My	3.0	1.5	0.0	3.0	2	0.0
2	Lady	0.0	3.0	3.0	2.5	3	0.0
3	Snakes	3.5	3.5	4.0	3.5	4	4.5
4	Superman	4.0	5.0	5.0	3.5	3	4.0
5	The Night	4.5	3.0	3.0	3.0	3	0.0
6	You Me	2.5	3.5	3.5	2.5	2	1.0

행렬
Itme x User

Toby 관람 영화(결측치 제외)

	title	Claudia	Gene	Jack	Lisa	Mick	Toby
3	Snakes	3.5	3.5	4.0	3.5	4	4.5
4	Superman	4.0	5.0	5.0	3.5	3	4.0
6	You Me	2.5	3.5	3.5	2.5	2	1.0

관람영화
평점

Toby와의 유사도(상관계수)

Claudia	Gene	Jack	Lisa	Mick
0.8934051	0.3812464	0.6628490	0.9912407	0.9244735

유사도 계산

movies

	critic	title	rating
1	Jack	Lady	3.0
2	Jack	Snakes	4.0
3	Jack	You Me	3.5
4	Jack	Superman	5.0
5	Jack	The Night	3.0
6	Mick	Lady	3.0
7	Mick	Snakes	4.0
8	Mick	Just My	2.0
9	Mick	Superman	3.0
10	Mick	You Me	2.0
11	Mick	The Night	3.0
12	Claudia	Snakes	3.5
13	Claudia	Just My	3.0
14	Claudia	You Me	2.5
15	Claudia	Superman	4.0
16	Claudia	The Night	4.5
17	Lisa	Lady	2.5
18	Lisa	Snakes	3.5
19	Lisa	Just My	3.0
20	Lisa	Superman	3.5
21	Lisa	The Night	3.0
22	Lisa	You Me	2.5
23	Toby	Snakes	4.5
24	Toby	Superman	4.0
25	Toby	You Me	1.0
26	Gene	Lady	3.0
27	Gene	Snakes	3.5
28	Gene	Just My	1.5
29	Gene	Superman	5.0
30	Gene	You Me	3.5
31	Gene	The Night	3.0

Toby의 미관람 영화 관측치 추출

	critic	title	rating
1	Jack	Lady	3.0
5	Jack	The Night	3.0
6	Mick	Lady	3.0
8	Mick	Just My	2.0
11	Mick	The Night	3.0
13	Claudia	Just My	3.0
16	Claudia	The Night	4.5
17	Lisa	Lady	2.5
19	Lisa	Just My	3.0
21	Lisa	The Night	3.0
26	Gene	Lady	3.0
28	Gene	Just My	1.5
31	Gene	The Night	3.0

Toby와의 유사도(상관계수)



	critic	similarity
Claudia	Claudia	0.8934051
Gene	Gene	0.3812464
Jack	Jack	0.6628490
Lisa	Lisa	0.9912407
Mick	Mick	0.9244735

유사도 포함

	critic	title	rating	similarity
1	Jack	Lady	3.0	0.6628490
2	Jack	The Night	3.0	0.6628490
3	Mick	Lady	3.0	0.9244735
4	Mick	Just My	2.0	0.9244735
5	Mick	The Night	3.0	0.9244735
6	Claudia	Just My	3.0	0.8934051
7	Claudia	The Night	4.5	0.8934051
8	Lisa	Lady	2.5	0.9912407
9	Lisa	Just My	3.0	0.9912407
10	Lisa	The Night	3.0	0.9912407
11	Gene	Lady	3.0	0.3812464
12	Gene	Just My	1.5	0.3812464
13	Gene	The Night	3.0	0.3812464

join

유사도 평점 계산 = Toby미관람 영화평점 * Toby와의 유사도

	critic	title	rating	similarity
1	Jack	Lady	3.0	0.6628490
2	Jack	The Night	3.0	0.6628490
3	Mick	Lady	3.0	0.9244735
4	Mick	Just My	2.0	0.9244735
5	Mick	The Night	3.0	0.9244735
6	Claudia	Just My	3.0	0.8934051
7	Claudia	The Night	4.5	0.8934051
8	Lisa	Lady	2.5	0.9912407
9	Lisa	Just My	3.0	0.9912407
10	Lisa	The Night	3.0	0.9912407
11	Gene	Lady	3.0	0.3812464
12	Gene	Just My	1.5	0.3812464
13	Gene	The Night	3.0	0.3812464



	critic	title	rating	similarity	sim_rating
1	Jack	Lady	3.0	0.6628490	1.9885469
2	Jack	The Night	3.0	0.6628490	1.9885469
3	Mick	Lady	3.0	0.9244735	2.7734204
4	Mick	Just My	2.0	0.9244735	1.8489469
5	Mick	The Night	3.0	0.9244735	2.7734204
6	Claudia	Just My	3.0	0.8934051	2.6802154
7	Claudia	The Night	4.5	0.8934051	4.0203232
8	Lisa	Lady	2.5	0.9912407	2.4781018
9	Lisa	Just My	3.0	0.9912407	2.9737221
10	Lisa	The Night	3.0	0.9912407	2.9737221
11	Gene	Lady	3.0	0.3812464	1.1437393
12	Gene	Just My	1.5	0.3812464	0.5718695
13	Gene	The Night	3.0	0.3812464	1.1437393

추천
영화

영화별 합계 계산

	title	similarity_tot	sim_rating_tot
1	Just My	3.190365732	8.074754106
2	Lady	2.959809565	8.383808341
3	The Night	3.853214712	12.899751858

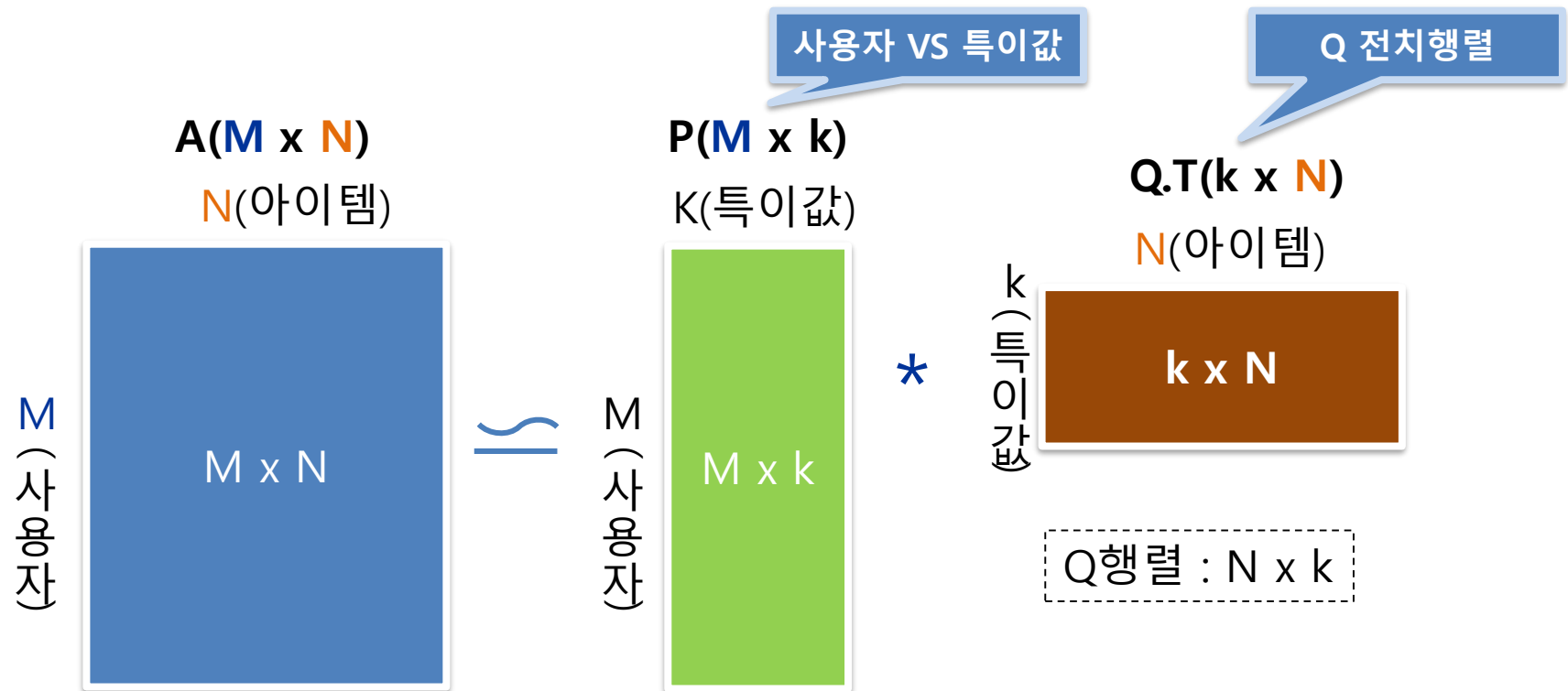


sim_rating_tot / similarity_tot

2.530980704
2.832549918
3.347789527

6) 특이값 분해(SVD)

- 특이값 행렬 분해(Singular Value Decomposition)
- 차원축소(dimension reduction) 기법 : 특이값 생성
- 행렬 $A(m \times n)$ 분해 예



SVD 특이값 예

특이값 : 행렬의 특징을 나타내는 값으로 알고리즘 이용
-> 차원축소에서 사용됨(5개 차원 -> 2개 차원 축소)

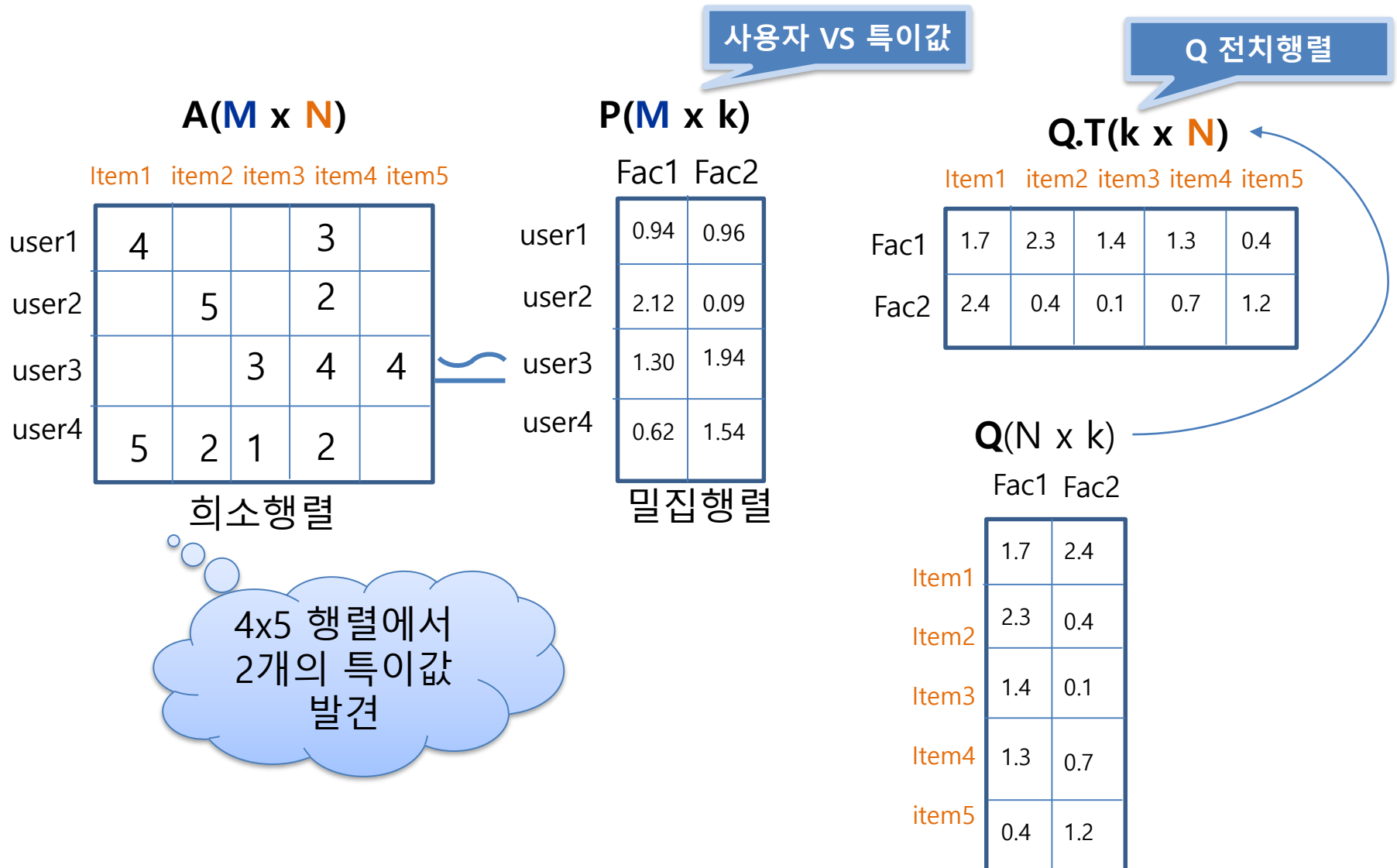
영화 평점 테이블

#	스타워즈	아바타	혹성탈출	사랑과영혼	타이타닉
# u1	1	1	1	0	0
# u2	3	3	3	0	0
# u3	4	4	4	0	0
# u4	5	5	5	0	0
# u5	0	0	0	4	4
# u6	0	0	0	5	5
# u7	0	0	0	2	2

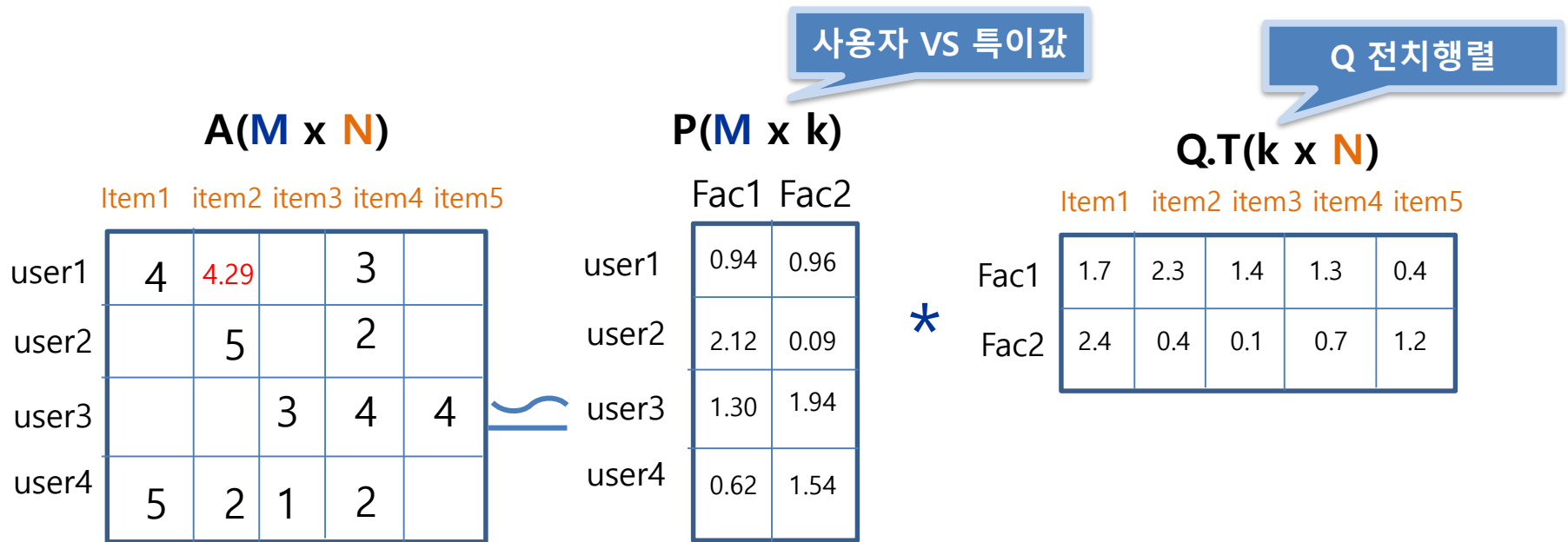
movieRating

#	스타워즈	아바타	혹성탈출	사랑과영혼	타이타닉
# u1	1	1	1	0	0
# u2	3	3	3	0	0
# u3	4	4	4	0	0
# u4	5	5	5	0	0
# u5	0	0	0	4	4
# u6	0	0	0	5	5
# u7	0	0	0	2	2

- 사용자 x 아이템 행렬 분해 예



- SVD 평점 예측 = 사용자 행렬 vs 아이템 행렬의 행렬곱



user1의 item1 평점 예측치

$$= 0.94 \times 1.7 + 0.96 \times 2.4$$

$$3.902 = 1.598 + 2.304$$

user1의 item2 평점 예측치

$$= 0.94 \times 2.3 + 0.96 \times 0.4$$

$$4.286 = 2.162 + 0.384$$

영화 평점 예측

2개의 대각 행렬 사용 영화 평점 예측

```
movieRating_svd3$u[,1:2]*%diag(movieRating_svd3$d[1:2])*%t(movieRating_svd3$v[,1:2])
```

##	[,1]	[,2]	[,3]	[,4]	[,5]
## [1,]	1.03	0.98	0.98	0.02	-0.03
## [2,]	3.09	2.95	2.95	0.07	-0.08
## [3,]	4.13	3.93	3.93	0.09	-0.10
## [4,]	5.16	4.92	4.92	0.11	-0.13
## [5,]	0.11	-0.06	-0.06	4.08	3.91
## [6,]	0.13	-0.07	-0.07	5.10	4.89
## [7,]	0.05	-0.03	-0.03	2.04	1.96
## [8,]	3.52	3.26	3.26	2.65	2.38

아바타 영화
추천

movieRating					
##	스타워즈	아바타	혹성탈출	사랑과영혼	타이타닉
## u1	1	1	1	0	0
## u2	3	3	3	0	0
## u3	4	4	4	0	0
## u4	5	5	5	0	0
## u5	0	0	0	4	4
## u6	0	0	0	5	5
## u7	0	0	0	2	2
## u8	4	3	3	3	2

- 영화 평점 예측 = 사용자 행렬 vs 아이템 행렬의 행렬곱

Surprise install

- Anaconda Prompt에서 설치

>conda install -c conda-forge scikit-surprise

```
Anaconda Prompt - conda install -c conda-forge scikit-surprise
environment location: C:\Anaconda3

added / updated specs:
- scikit-surprise

The following packages will be downloaded:

package                                build                                size  channel
-----                                -
scikit-surprise-1.0.6                  py36h452e1ab_1                    589 KB  conda-forge
certifi-2018.4.16                      py36_0                            143 KB  conda-forge
joblib-0.12.5                          py36_0                            176 KB  conda-forge
conda-4.5.11                          py36_0                            654 KB  conda-forge

Total: 1.5 MB

The following NEW packages will be INSTALLED:

joblib: 0.12.5-py_0 conda-forge
scikit-surprise: 1.0.6-py36h452e1ab_1 conda-forge

The following packages will be UPDATED:

certifi: 2018.4.16-py36_0 conda-forge
conda: 4.5.9-py36_0 conda-forge

Proceed ([y]/n)? y
```



```
Anaconda Prompt

joblib-0.12.5                          py_0                            176 KB  conda-forge
conda-4.5.11                          py36_0                            654 KB  conda-forge

Total: 1.5 MB

The following NEW packages will be INSTALLED:

joblib: 0.12.5-py_0 conda-forge
scikit-surprise: 1.0.6-py36h452e1ab_1 conda-forge

The following packages will be UPDATED:

certifi: 2018.4.16-py36_0 conda-forge
conda: 4.5.9-py36_0 conda-forge

Proceed ([y]/n)? y

Downloading and Extracting Packages
scikit-surprise-1.0.6-py36h452e1ab_1 589 KB |#####| 100%
certifi-2018.4.16-py36_0 143 KB |#####| 100%
joblib-0.12.5-py_0 176 KB |#####| 100%
conda-4.5.11-py36_0 654 KB |#####| 100%
Preparing transaction: done
Verifying transaction: done
Executing transaction: done

(base) C:\Users\python>
```