

# 데이터 분석 보고서

단골 브랜드 데이터 활용 유사도 기반 추천 시스템

1. 개요
  - . 현 상황 진단
  - . 선행 연구 탐색 및 연구가설 설정
2. 분석 결과
  - . 추천 알고리즘 요약
  - . 데이터 현황
  - . 데이터 분석 결과
  - . 성능 평가
3. 결론

# 개요

## 1. 현 상황 진단

〈‘상품이 적다’ VOC 지속 접수〉

바이어(특히 해외바이어)로부터  
**‘상품(SKU)가 적다’** 라는 VoC가 **지속적으로** 접수

〈검색어 결과 품질 이슈〉

바이어(특히 해외 바이어)의 경우 **검색 결과를 신뢰할 수 없어**  
**카테고리별 상품 탐색**을 진행하는 경향이 있음

〈C레벨 희망 메인화면 개편 방향〉



카테고리를 넘어 **주제별 종합 큐레이션**이 가능한 **메인화면**이 구성되었으면 하는 방향성을 제시

목표 설정

카테고리를 넘어 **나와 비슷한 다른 바이어의 연관 상품**이 **나에게도 노출되도록 추천 알고리즘 구성** 및 제공

## 2. 선행 연구 검토 및 연구 가설 설정

### 1) 선행 연구 탐색

#### - 콘텐츠 기반 필터링

- ① **아이템** 자체의 **특성**을 이용하여 **유사한 상품**을 추천하는 방법론
- ② 가령, 영화라고 한다면 영화에 대한 **‘장르’, ‘감독명’, ‘출연 배우’** 등이 **영화 A**라는 콘텐츠의 **특성**으로 작용
- ③ 콘텐츠 기반 필터링은 다음과 같은 **세 단계(부분)**으로 **일반화**가 가능함(S.M Mahdi Seyednezhad et.al, 2019)
  - **Content Analyzer** : 콘텐츠의 특성을 분석하는 부분. Raw데이터를 정형화된 정보로 변환한다
  - **Profile Learner** : Content Analyzer에서 추출한 콘텐츠의 특성을 추천 목표인 유저에 맞게 재가공하는 부분
  - **Filtering Component** : Profile Learner에서 나온 user Profile 기반으로 가장 유사한 상품을 추천하는 부분. 다양한 유사도 측정 방법론 활용 가능

		블록버스터	아동	SF	Over 1 million	
선택 영화		Yes	No	No	Yes	유사도 높음
		No	No	Yes	Yes	
		No	No	No	No	
		Yes	No	No	Yes	
추천 영화						

<콘텐츠 기반 필터링 예시>

## 2. 선행 연구 검토 및 연구 가설 설정

### 1) 선행 연구 탐색

#### - 협업 필터링

- ① 같은 영역에 존재하는 다른 공간의 데이터를 이용하여 추천 사항을 추론하는 방법론
- ② 가령, '유저의 구매 내역'이라는 같은 영역에 존재하는 '유저A'와 '유저B'의 경우, 둘이 충분히 유사하다면 A에게 B의 내역을 추천 가능
- ③ 협업 기반 필터링은 아래와 같이 분류 가능(S.M Mahdi Seyednezhad et.al, 2019)
  - 메모리 기반 협업 필터링 : User - Item 행렬에서만 작동. 추천 직전에 모든 평점을 직접적으로 업데이트하여 활용(배치성 활용)
  - 모델 기반 협업 필터링 : 학습 가능한 파라미터를 가진 모델을 활용하여 학습한 패턴을 기반으로 추천을 수행 (Paul Convington, Jay Adams and Emre Sargin, 2016)

					
유저 A	Yes	No	Yes	?	Yes 예측
유저 B	No	No	Yes	Yes	
유저 C	No	No	No	No	
유저 D	Yes	No	Yes	Yes	

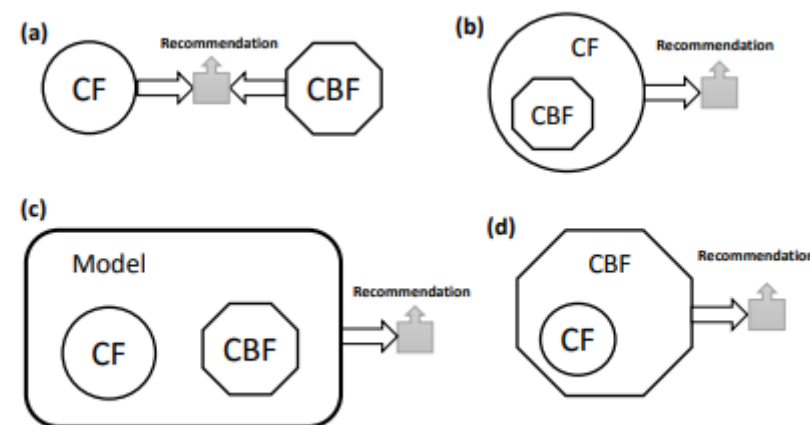
<협업 필터링 예시>

## 2. 선행 연구 검토 및 연구 가설 설정

### 1) 선행 연구 탐색

#### - 하이브리드 방법론

- ① 콘텐츠 기반 필터링과 협업 기반 필터링은 각자 고유한 장단점을 보유
- ② 주류 추천시스템은 협업 기반 필터링과 콘텐츠 기반 필터링을 결합(Hybrid)하여 추천을 수행(Keunho Choi et.al)
- ③ 우측 그림에 대한 설명은 아래와 같음
  - (a)의 경우 : 협업 필터링과 콘텐츠 기반 필터링의 결과물을 가중치를 통해 결합
  - (b)의 경우 : 콘텐츠 기반 필터링을 통해 도출된 콘텐츠 특성을 기반으로 협업 기반 필터링이 유저간 유사도를 측정
  - (c)의 경우 : 협업 필터링과 콘텐츠 기반 필터링의 Output을 또다른 예측 모델에 Input으로 투입
  - (d)의 경우 : 콘텐츠 기반 필터링의 User Profile 과정에 협업 필터링의 결과물을 활용하여 품질을 향상



<하이브리드 방식 개념도, S.M Mahdi Seyednezhad et.al, 2019 >

## 2. 선행 연구 검토 및 연구 가설 설정

### 2) 가설 설정

-. 가설 1 : 하이브리드 방법론 (d)를 차용한 추천 시스템은 단순 user-user 협업 필터링에 비해 실제 서비스에 활용 가능한 준수한 성능을 보일것이다

-. 가설 2 : 유저의 단골브랜드 추가 내역, 위시리스트 상품 추가 내역, 구매 내역을 모두 활용한 모델은 준수한 성능을 보일 것이다



본문

## 1. 추천 알고리즘 요약

1) 추천 A-1안	2) 추천 A-3안
<p><b>- 공통점</b></p> <p>① <b>유저 - 유저 협업 필터링</b>인 ‘유저 프로파일링’단계 존재 : <b>단골브랜드</b> 기반으로 유저간 유사도가 가장 높은 <b>Top - K (혹은 &gt; threshold)</b> 바이어를 후보 바이어로 추출</p> <p>② <b>공통 카테고리 추출</b> 단계 존재 : <b>Target Buyer</b>의 <b>상품 카테고리</b>와 일치하는 후보 바이어들의 상품들을 <b>후보 상품</b>으로 추출</p> <p>③ <b>상품 컨텐츠 기반 필터링</b> 존재 : 상품의 특성(<b>위시리스트를 추가한 바이어</b>)을 기반으로 <b>Target Buyer</b>의 <b>상품별</b>로 가장 유사도가 높은 후보 상품 <b>Top - K (혹은 &gt; Threshold)</b> 상품들을 최종 도출</p>	
<p><b>- 차이점(A-1안) : 구매 내역</b></p> <p>① Target Buyer의 상품과 후보 상품을 ‘공통 카테고리 추출 단계’, ‘상품 컨텐츠 기반 필터링’ 단계에서 추출 / 계산할 때 각 바이어의 <b>구매 내역</b> 기반으로 동작</p>	<p><b>- 차이점(A-3안) : 위시리스트</b></p> <p>① Target Buyer의 상품과 후보 상품을 ‘공통 카테고리 추출 단계’, ‘상품 컨텐츠 기반 필터링’ 단계에서 추출 / 계산할 때 각 바이어의 <b>위시리스트 추가 내역</b> 기반으로 동작</p>

VS

## 2. 데이터 현황

### 1) 데이터셋 정의

번호	컬럼명	타입	설명
1	brandId	Integer	브랜드ID
2	BuyerId	Integer	바이어ID
3	Yearmon	Datetime	추가일자
4	Value	Integer	내부 처리용(1 고정)

〈단골브랜드 추가 내역〉

번호	컬럼명	타입	설명
1	brandId	Integer	브랜드ID
2	BuyerId	Integer	바이어ID
3	productId	Integer	상품ID
4	initQuantity	Integer	내부 처리용(1 고정)
5	Status	Integer	주문 상태
6	createdAt	Datetime	주문 일자

〈주문 내역〉

번호	컬럼명	타입	설명
1	brandId	Integer	브랜드ID
2	buyerId	Integer	바이어ID
3	productId	Integer	상품ID
4	createdAt	Integer	추가일자

〈위시리스트 추가 내역〉

번호	컬럼명	타입	설명
1	productId	Integer	브랜드ID
2	name	String	상품명
3	Category	Array	상품 카테고리 목록
4	Images	Json	상품 이미지 목록

〈상품 목록〉

## 3. 데이터 분석 결과

### 1) 추천 A-1안

#### - 유저 프로파일링

- ① 바이어의 **단골 브랜드 추가 내역**을 기반으로 **유사한 바이어**를 도출하는 단계
- ② 각 바이어별로 순회를 돌면서 **Top-K or 역치(Threshold)** 이상의 **유사 브랜드**를 **코사인 유사도** 기반으로 추출
- ③ 해당 바이어들은 '**유사한 디자인 특성을 선호한다**' 라는 가정하에 다음 절차를 진행

1. 유저  
프로파일링

2. 동일  
카테고리  
도출

3. 상품  
프로파일링

브랜드ID	2123	5211	33321	18620
바이어ID				
1215	1	0	0	1
1224	1	1	0	1
1351	0	0	0	0
1455	1	1	1	1
3851	0	0	0	1

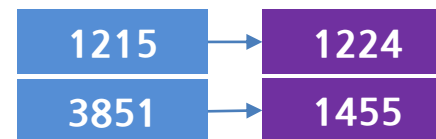
<유저 - 브랜드 협업 행렬>

유사도 계산

	1215	1224	1351	1455	3851
1215	1	0.82	0	0.71	0.71
1224	0.82	1	0	0.86	0.58
1351	0	0	0	0	0
1455	0.70	0.86	0	1	0.5
3851	0.	0.5	0	0.58	1

<유저간 코사인 유사도 행렬>

Top-K 추출 (K = 1 예시)



## 3. 데이터 분석 결과

### 1) 추천 A-1안

#### - 동일 카테고리 추출

① 각 바이어별로 순회를 돌면서 바이어가 구매한 각 상품별로 3-Depth까지 동일한 유사 바이어의 상품들을 도출

1. 유저  
프로파일링

2. 동일  
카테고리  
도출

3. 상품  
프로파일링

<1215 바이어의 주문내역>

상품ID	카테고리
1278542	아동복 / 하의 / 바지
1385521	여성복 / 드레스 / 드레스
3715215	여성복 / 상의 / 셔츠
1522182	남성복 / 외투 / 카디건
5221852	남성복 / 상의 / 셔츠
3521852	악세서리 / 가방 / 클러치

<1224 바이어의 주문내역>

상품ID	카테고리
1255258	여성복 / 드레스 / 드레스
2352218	여성복 / 상의 / 셔츠
1152851	남성복 / 외투 / 카디건
5221852	남성복 / 상의 / 티셔츠
3215582	악세서리 / 패션 / 귀걸이
1212153	악세서리 / 가죽 / 벨트

상품 도출

1255258

1152851

...

## 3. 데이터 분석 결과

### 1) 추천 A-1안

#### - 상품 프로파일링

① 도출된 동일 카테고리내 유사 상품들 대상

② 상품 특성들을 이용하여 Top - K의 상품들을 유사도 추출

. 현재는 상품을 추가한 바이어의 위시리스트를 상품 특성으로 활용  
. 향후 상품 특성이 더 뚜렷하게 드러나는 다른 특성(이미지, 텍스트 등)으로 교체 가능

③ 추출한 상품들을 유사도기준으로 Sorting하여 최종 표출

1. 유저  
프로파일링

2. 동일  
카테고리  
도출

3. 상품  
프로파일링

바이어ID	2123	5211	33321	18620	...
상품ID					...
1278542	1	0	0	1	
1385521	1	1	0	1	
3715215	0	0	0	0	
1255258	1	1	1	1	
1152851	0	0	0	1	
...					

〈상품 특성 행렬(위시리스트)〉

유사도 계산

	1255258	1152851	...
1278542	0.23	0.33	...
1385521	0.10	0.12	
3715215	0.08	0.02	

리스트 도출

바이어	상품ID	상품명	유사도
1226	1255258	니트 스커트	0.33
	1152851	신상 남성 카디건	0.12

## 3. 데이터 분석 결과

### 2) 추천 A-3안

#### - 유저 프로파일링

- ① 바이어의 **단골 브랜드 추가 내역**을 기반으로 **유사한 바이어**를 도출하는 단계
- ② 각 바이어별로 순회를 돌면서 **Top-K or 역치(Threshold)** 이상의 **유사 브랜드**를 **코사인 유사도** 기반으로 추출
- ③ 해당 바이어들은 '**유사한 디자인 특성을 선호한다**' 라는 가정하에 다음 절차를 진행

1. 유저  
프로파일링

2. 동일  
카테고리  
도출

3. 상품  
프로파일링

브랜드ID	2123	5211	33321	18620
바이어ID				
1215	1	0	0	1
1224	1	1	0	1
1351	0	0	0	0
1455	1	1	1	1
3851	0	0	0	1

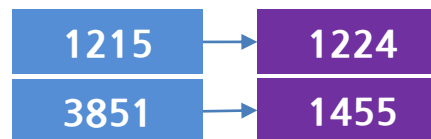
<유저 - 브랜드 협업 행렬>

유사도 계산

	1215	1224	1351	1455	3851
1215	1	0.82	0	0.71	0.71
1224	0.82	1	0	0.86	0.58
1351	0	0	0	0	0
1455	0.70	0.86	0	1	0.5
3851	0.	0.5	0	0.58	1

<유저간 코사인 유사도 행렬>

Top-K 추출 (K = 1 예시)



## 3. 데이터 분석 결과

### 2) 추천 A-3안

#### - 동일 카테고리 추출

- ① 각 바이어별로 순회를 돌면서 바이어가 위시리스트로 추가한 각 상품들의 카테고리 도출
- ② 유저 프로파일링 단계에서 도출한 후보 바이어들의 위시리스트 추가 상품들의 카테고리 도출
- ③ 3-Depth까지 동일한 유사 바이어의 상품들을 도출

<1215 바이어의 위시리스트>

상품ID	카테고리
1278542	아동복 / 하의 / 바지
1385521	여성복 / 드레스 / 드레스
3715215	여성복 / 상의 / 셔츠
1522182	남성복 / 외투 / 카디건
5221852	남성복 / 상의 / 셔츠
3521852	악세서리 / 가방 / 클러치

<1224 바이어의 위시리스트>

상품ID	카테고리
1255258	여성복 / 드레스 / 드레스
2352218	여성복 / 상의 / 셔츠
1152851	남성복 / 외투 / 카디건
5221852	남성복 / 상의 / 티셔츠
3215582	악세서리 / 패션 / 귀걸이
1212153	악세서리 / 가죽 / 벨트

#### 상품 도출

1255258  
1152851  
...

1. 유저  
프로파일링

2. 동일  
카테고리  
도출

3. 상품  
프로파일링



## 3. 데이터 분석 결과

### 2) 추천 A-3안

#### - 상품 프로파일링

① 도출된 동일 카테고리내 유사 상품들 대상

② 상품 특성들을 이용하여 Top - K 유저의 위시리스트 상품들의 유사도 추출

. 현재는 상품을 추가한 바이어의 위시리스트를 상품 특성으로 다시 활용  
 . 향후 상품 특성이 더 뚜렷하게 드러나는 다른 특성(이미지, 텍스트 등)으로 교체 가능

③ 추출한 상품들을 유사도기준으로 Sorting하여 최종 표출

1. 유저  
프로파일링

2. 동일  
카테고리  
도출

3. 상품  
프로파일링

바이어ID	2123	5211	33321	18620	...
상품ID					...
1278542	1	0	0	1	
1385521	1	1	0	1	
3715215	0	0	0	0	
1255258	1	1	1	1	
1152851	0	0	0	1	
...					

〈상품 특성 행렬(위시리스트)〉

유사도 계산

	1255258	1152851	...
1278542	0.23	0.33	...
1385521	0.10	0.12	
3715215	0.08	0.02	

리스트 도출

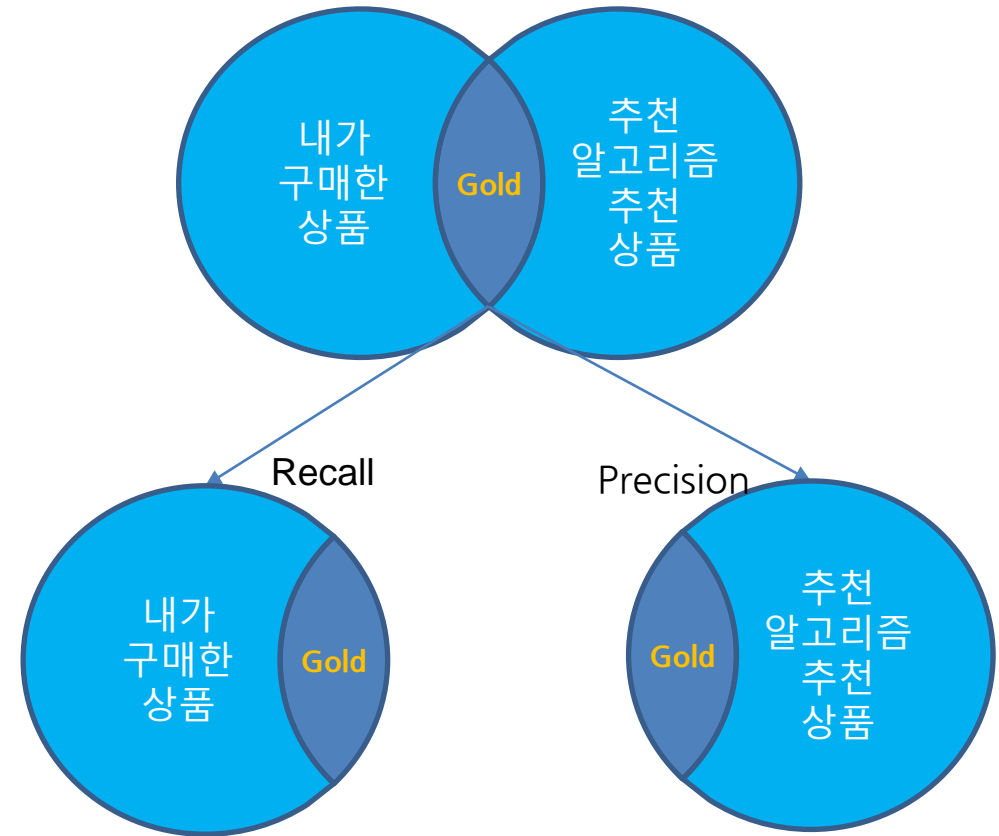
바이어	상품ID	상품명	유사도
1226	1255258	니트 스커트	0.33
	1152851	신상 남성 카디건	0.12

## 4. 성능 평가

### 1) 성능 평가 지표

#### - 성능 측정 결과

- ① 내가 실제 구매한 상품 중 추천 알고리즘이 추천한 상품의 비율 -> Recall(재현율)
- ② 내가 구매했다고 추천한 상품 중 실제 내가 구매한 상품의 비율 -> Precision(정밀도)



## 4. 성능 평가

### 2) 추천 A-1안 성능 측정 결과

#### - 성능 측정 결과

- ① 내가 실제 구매한 상품 중 추천 알고리즘이 추천한 상품의 비율 -> Recall(재현율)
- ② 내가 구매했다고 추천한 상품 중 실제 내가 구매한 상품의 비율 -> Precision(정밀도)
- ③ 추천 알고리즘을 가동하면서 데이터 부족으로 전처리 불가 / 타겟 상품 도출 불가 / 후보 상품 도출 불가 / 추천 결과 도출 불가로 제외된 빈도

지표	값
Precision	0.012
Recall	0.029

<A-1안 성능 평가표>

지표	값	
총 명수	6,446명	
전처리 후	4,641명	
추천 도출 제외	타겟 상품 X	19
	후보 상품 X	0
	추천 결과 X	691
	위시리스트 X	0
추천 수행	3,931명	

<A-1안 콜드 스타트 사례 빈도표(23.03.07 기준)>

## 4. 성능 평가

### 3) 추천 A-3안 성능 측정 결과

#### - 성능 측정 결과

- ① 내가 실제 구매한 상품 중 추천 알고리즘이 추천한 상품의 비율 -> Recall(재현율)
- ② 내가 구매했다고 추천한 상품 중 실제 내가 구매한 상품의 비율 -> Precision(정밀도)
- ③ 추천 알고리즘을 가동하면서 데이터 부족으로 전처리 불가 / 타겟 상품 도출 불가 / 후보 상품 도출 불가 / 추천 결과 도출 불가로 제외된 빈도

지표	값
Precision	0.014
Recall	0.046

〈A-3안 성능 평가표〉

지표	값	
총 명수	6,446명	
전처리 후	4,641명	
추천 도출 제외	타겟 상품 X	165
	후보 상품 X	34
	추천 결과 X	684
	위시리스트 X	16
추천 수행	3,740명	

〈A-3안 콜드 스타트 사례 빈도표(23.03.07 기준)〉

## 4. 성능 평가

### 4) 비교군 : user - user 협업필터링

#### - 성능 측정 결과

① A-1, A-3안과 비교했을 때 성능지표와 콜드스타트 양쪽 모두 USER - USER 협업 필터링이 우수한 것으로 언뜻 보임

② 그러나, 실제 사용자 측면에서 추천 결과를 확인해봤을 때 다음의 절차 결여로 매력적인 추천 결과를 뽑아내지 못하는 것으로 보임

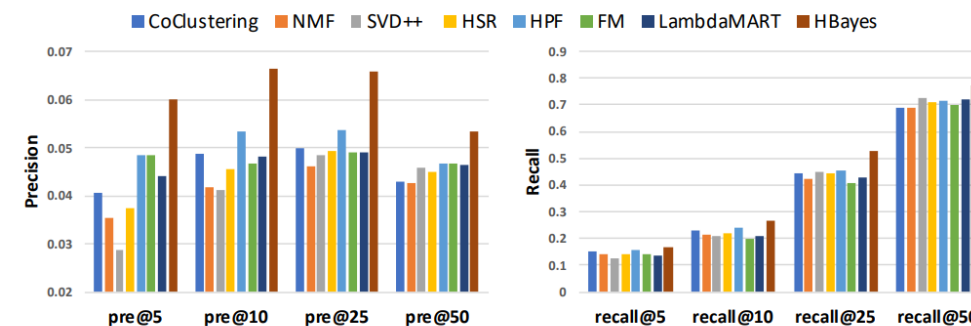
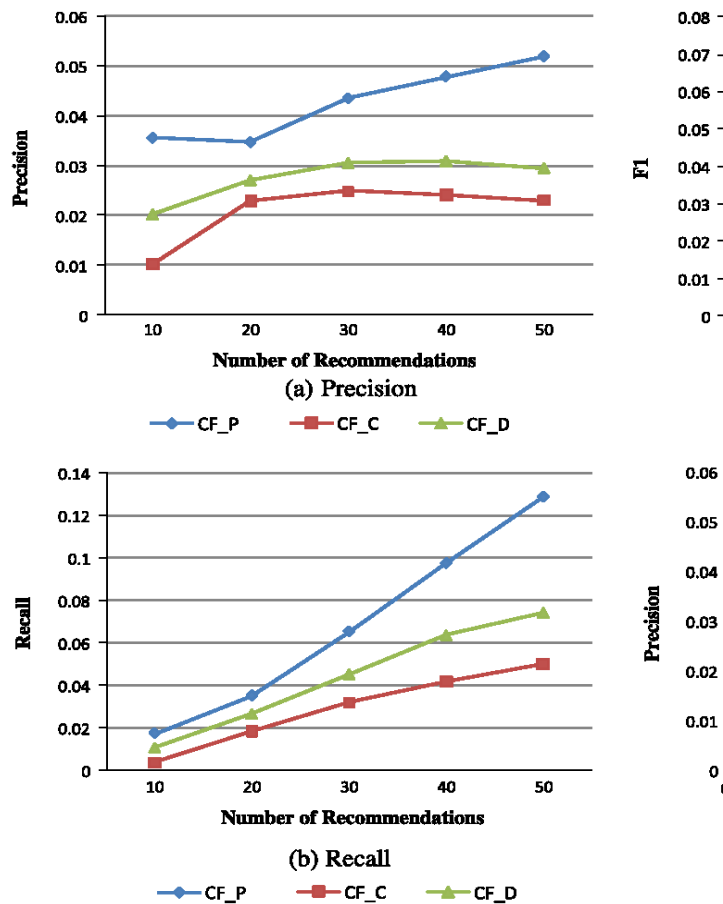
- **카테고리 기반 필터링 부재** : Target-buyer의 구매 상품별 **카테고리를 고려 X** → Target-buyer가 **별로 구매하지 않은 카테고리의 상품을 다수 추천**
- **추천 결과 부재 多** : A-1, A-3안 모두 **추천결과가 없는 경우가 7~800건 수준**이나, A-3안 같은 경우는 **1778건으로 2.5배 수준**  
→ A-1, A-3안의 경우 **wishlist, 단골브랜드 내역 추가**를 통해 **Cold-start 해결**이 가능하나 이 방법의 경우 **유저가 구매를 더 많이** 하지 않는 한 해결 불가능

지표	값
Precision	0.287
Recall	0.11

〈협업 필터링 성능 평가표〉

지표	값	
총 명수	6,446명	
전처리 후	-	
추천 도출 제외	타겟 상품 X	-
	후보 상품 X	-
	추천 결과 X	1778
	위시리스트 X	-
추천 수행	4,668명	

〈협업 필터링콜드 스타트 사례 빈도표(23.03.07 기준)〉



A hybrid online-product recommendation system: Combining implicit rating-based collaborative filtering and sequential pattern analysis

Hierarchical Bayesian Personalized Recommendation: A Case Study and Beyond

# 결론

## 1. 결론 요약

### 결론 1

- 정량적 지표로만 확인했을 때 단순 user-user 협업필터링이 A-1, A-3안보다 더 우수
- But 정성적으로 확인했을 때 다음 사유로 A-1, A-3이 더 우수한 것으로 판단
  - ① 추천 적정성 : 사용자 경험 차원에서 A-1, A-3이 user-user 협업 필터링보다 더 적절한 결과를 추천
  - ② 성능 개선 가능성 : 추천 실패 사례수가 2.5배에 이르고 추가적인 개선이 불가능하여 A-1, A-3안이 더 우수한 것으로 보임

### 결론 2

- 정량적 지표로 확인했을 때 Precision / Recall은 0.02 / 0.04 수준으로 유사 연구들과 비교해봤을 때 평균적인 성능으로 판단
- 정성적으로 확인했을 때 유사한 스타일과 같은 카테고리를 기반으로 추천을 수행하기 때문에 사용자 경험 차원에서 적절한 추천 결과를 보여주는 것으로 판단



## 2. 연구의 한계 향후 시사점

### 연구의 한계

- 연구 환경의 한계 : GPU 활용이 불가능하고, 메모리가 작은 현 연구 환경에서 좀 더 장기간 / 좀 더 많은 변수를 활용 / Matrix Factorization같은 좀 더 상호간의 연결성을 강조하는 방법론 활용에 한계

### 향후 시사점

- 향후 연구 환경이 개선되고, 유저들이 Wishlist / 단골 브랜드 추가를 더 적극적으로 수행할 경우 추천 성능과 Cold-start 문제 해결에 진전이 있을 것으로 기대

-

- S.M Mahdi Seyednezhad et.al, “A Review on Recommendation Systems: Context-aware to Social-based”, Arxiv(2018),
- Zitao Liu Et.al, “Hierarchical Bayesian Personalized Recommendation”, Arxiv(2019)
- Keunho Choi et.al, “A hybrid online-product recommendation system: Combining implicit rating-based collaborative filtering and sequential pattern analysis”, Electronic Commerce Research and Applications Vol 11, Issue 4(2012), pp 309-317
- Paul Convington, Jay Adams and Emre Sargin, “Deep Neural Networks for YouTube Recommendations”, RecSys '16: Proceedings of the 10th ACM Conference on Recommender Systems(2016), pp 191-198